

VOLUME 104

PART C NUMBER 5

MARCH 1957



The Proceedings
OF
THE INSTITUTION OF
ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C
MONOGRAPHS Nos. 187-211

SAVOY PLACE · LONDON W.C.2

Price Twelve Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871

INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1956-1957

President

SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).

Past-Presidents

SIR JAMES SWINBURNE, Bart., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGECUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C.,
T.D., D.Sc.(Eng.).

SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
J. ECCLES, C.B.E., D.Sc.
SIR GEORGE H. NELSON, Bart.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.

G. S. C. LUCAS, O.B.E.
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D.,
B.Sc.

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH.

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
J. A. BROUGHALL, B.Sc.(Eng.).
C. M. COCK.
A. R. COOPER, M.Eng.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
E. M. HICKIN.
F. J. LANE, O.B.E., M.Sc.
D. McDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.

H. H. MULLENS, B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
D. P. SAYERS, B.Sc.
G. L. WATES, J.P.
H. WATSON-JONES, M.Eng.
D. B. WELBOURN, M.A.
H. WEST, M.Sc.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.

Chairmen and Past-Chairmen of Sections

Measurement and Control:

D. TAYLOR, M.Sc., Ph.D.
*W. BAMFORD, B.Sc.

Radio and Telecommunication:

R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
*H. STANESBY.

Supply:

P. J. RYLE, B.Sc.(Eng.).
*L. DRUCQUER.

Utilization:

H. J. GIBSON, B.Sc.
*D. B. HOGG, M.B.E., T.D.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:

H. L. HASLEGRAVE, M.A., Ph.D., M.Sc.
(Eng.).
*F. R. C. ROBERTS.

Mersey and North Wales Centre:

P. D'E. STOWELL, B.Sc.
*PROFESSOR J. M. MEEK, D.Eng.

North Midland Centre:

W. K. FLEMING.
*F. BARRELL.

North-Eastern Centre:

*J. CHRISTIE.
A. H. KENYON.

North-Western Centre:

T. E. DANIEL, M.Eng.
*G. V. SADLER.

Northern Ireland Centre:

DOUGLAS S. PARRY.
*MAJOR E. N. CUNLIFFE, B.Sc.Tech.

Western Centre:

PROFESSOR G. H. RAWCLIFFE, M.A., D.Sc.
*T. G. DASH, J.P.

* Past-Chairman.

Scottish Centre:

PROFESSOR F. M. BRUCE, M.Sc., Ph.D.
*E. WILKINSON, Ph.D., B.Eng.

South Midland Centre:

C. J. O. GARRARD, M.Sc.
*H. S. DAVIDSON, T.D.

Southern Centre:

H. ROBSON, B.Sc.
*L. H. FULLER, B.Sc.(Eng.).

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers. An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 104. PART C. No. 5.

MARCH 1957

DISCUSSION ON

'PEAK-VOLTAGE MEASUREMENTS OF STANDARD IMPULSE VOLTAGE WAVES'*

Mr. G. C. Dewsnap (*communicated*): In determining the design requirements of a recurrent surge generator and its associated oscilloscope, I had cause to obtain the constants and frequency spectrum of a standard 1/50 microsec wave. The constants so determined agree with those quoted in the paper. Similarly, the requirements for the frequency response of equipment used to view the wave were obtained by specifying ability to pass frequencies of an amplitude down to 1% of the maximum amplitude in the wave. As a result for the 1/50 microsec wave, the upper limit of frequency was found to be about 220 kc/s.

The method adopted by Mr. Aked of specifying frequencies obtained from the shorter time-constant of the wave equation is surely far too exacting. What in effect has been done has been to take the frequency spectrum of an impulse of infinitely short rise time and of decay time-constant of $1/(6.073 \times 10^6)$ sec and equate it to the frequency spectrum of the surge, which has a rise time of the order of 1 microsec and a considerably longer decay time. The two terms that make up the equation to the wave surely should be taken together, and if this is done the frequency spectrum of the first term of eqn. (2) is found to be the predominant one. This term still represents a function which has an infinitely short rise time and a tail which approximately corresponds to the decay of the actual surge. The sharp rise of this function provides a more exacting condition to be satisfied than the complete function; i.e. its frequency spectrum contains significant terms of higher frequency than the complete function itself, so that a frequency spectrum up to 226 kc/s is more than adequate in specifying the required response of any associated equipment.

A further error which completely invalidates the frequency response tests carried out on the divider has been made in determining the upper limit of R_{a2} , which (from a correct evaluation of the author's own figures) should be 96.7 Mc/s, not 9.67 Mc/s.

If the author's arguments were correct the divider should have

been tested to 30 Mc/s, which would be fatal for the operation of equipment far more complex than that described.

The equipment designed and described should prove quite satisfactory for its purpose, but certainly not because of the theoretical analysis accompanying it.

Mr. A. Aked (*in reply*): As Mr. Dewsnap observes, the upper limit of R_{a2} should be 96.7 Mc/s and not 9.67 Mc/s as stated in the paper. This error makes three alterations necessary in the text; 96.7 Mc/s must be substituted for 9.67 Mc/s in Section 4, par. 3, and 100 Mc/s must be substituted for 10 Mc/s in Section 4, par. 4, and in Section 5, sentence 1.

Eqn. (2) represents the ideal impulse wave, and any circuit required to record this wave faithfully would have to have a level response up to nearly 100 Mc/s. This is the frequency at which the amplitude of the spectrum of the wavefront component has fallen to 1% of its maximum value. The recording of this component is essential if the wavefront is to be recorded faithfully.

The paper, however, was concerned with the measurement of the peak value of the wave, and therefore the range of frequencies over which the divider response had to be known was reduced. To determine the limits of this frequency range, the error produced by neglecting the components in the spectrum above a given frequency was determined for each of the components of the ideal wave. As stated in the paper, for a cut-off frequency of 1 Mc/s the estimated value of the wavetail component at 1 microsec differed by only 0.16% from its true value.

At 10 Mc/s the amplitude of the spectrum of the wavetail exponential is 9.6% of its maximum. If the components above 10 Mc/s in this spectrum are neglected, the estimated value of the exponential at 1 microsec is 0.23% of its maximum while the true value is 0.16%.

Even with the high-frequency components of the ideal wave, only relatively small errors would be introduced by not knowing the divider response above 10 Mc/s.

Section 11 explains how possible errors due to the peaking of the divider response above 3 Mc/s were shown to be negligible for the actual impulse waves used.

* AKED, A.: Monograph No. 153 M, November, 1955 (see 103 C, p. 186).

AN INVESTIGATION OF THE EDDY-CURRENT ANOMALY IN A LOW-SILICON SHEET STEEL

By H. ASPDEN, Ph.D., B.Sc., Wh.Sc., Associate Member.

(The paper was first received 9th February, and in revised form 16th April, 1956. It was published as an INSTITUTION MONOGRAPH in July, 1956.)

SUMMARY

The eddy-current anomaly effects in an 0.019 in thick low-silicon sheet steel are investigated and analysed on the basis of the magnetic inhomogeneity arising from ferromagnetic domain structure. This particular steel was chosen because previous experimental work had shown that inherent time-lag effects in the magnetization process were small. However, in such a steel, flux waveform distortion arising from hysteresis phenomena adds to the anomaly effect, and a special experimental technique by which such distortion effects could be eliminated from the analysis was developed. This made it possible to test the steel, focusing attention upon the anomaly effect attributable to the magnetic inhomogeneity. The results obtained are interpreted in terms of a hypothetical domain configuration, and it is concluded that the magnetic inhomogeneity arising from ferromagnetic domain structure does account for the eddy-current anomaly, distortion and time-lag effects being of minor importance.

LIST OF PRINCIPAL SYMBOLS

- B = Flux density, gauss.
 B_m = Amplitude of flux density, gauss.
 H = Magnetizing field strength, oersteds.
 H_d = Demagnetizing field strength, oersteds.
 b = Half thickness of sheet steel, cm.
 ω = Angular frequency of magnetization, rad/sec.
 t = Time, sec.
 ρ = Resistivity of sheet material, ohm-cm.
 θ = Phase angle between maxima of B and H , having a theoretical value of $\frac{4\pi}{3} \frac{b^2 \omega \mu}{\rho \times 10^9}$ rad.
 μ = Apparent differential permeability.
 μ_a = Actual differential permeability.
 μ_m = Maximum differential permeability.

(1) INTRODUCTION

When the eddy-current loss in an electrical sheet steel magnetized at a power frequency is measured, it is invariably found to exceed that calculated from classical formulae. These formulae are based upon an assumed magnetic homogeneity, and this assumption is usually justified by ensuring that the sheet steel has been very well annealed. The anomaly is then accounted for in terms of effects attributable to flux waveform distortion arising from the non-linear character of magnetic hysteresis.¹ However, in recent years it has become apparent that in sheet steel the actual magnetic domain configuration does not support the assumption that the steel is magnetically homogeneous, and that this can lead to high anomaly factors.² It has also been suggested that a time-lag effect in the magnetization process may well exist.³ This would lead to an extra loss having the same character as eddy-current loss and would therefore add to the anomaly effect.

This recent work has left open the question whether flux waveform distortion contributes appreciably to the discrepancy

loss associated with the eddy-current anomaly. This is an important question, since distortion effects arising from the well-known non-linear form of the hysteresis loop afford a most plausible qualitative explanation and therefore provide a generally accepted explanation for the anomaly.

It is the object of the paper to consider the importance of such distortion effects. It will be argued that they are much less important than is generally believed, and to show this, a relatively-thick low-resistivity sheet steel, in which distortion effects should be relatively more important, was tested by a novel experimental technique by which distortion effects can be avoided and the measurements directed solely to an investigation of the anomaly as it arises due to other causes.

This particular steel is one in which evidence from other experimental work shows that the intrinsic time-lag effect in the magnetization process is small.² It would appear, therefore, that, by eliminating distortion effects, it is possible to study the anomaly as it arises almost wholly from magnetic inhomogeneity. A second and principal object of this work is therefore to analyse the results obtained with a view to assessing the nature of this inhomogeneity.

A preliminary task is to assess the relative importance of the effects of flux waveform distortion in enhancing the apparent discrepancy between actual and theoretical eddy-current effects.

(2) FLUX-WAVEFORM-DISTORTION EFFECTS

(2.1) Brailsford's Work on Flux Waveform Distortion

Brailsford¹ has measured an actual loss anomaly factor of 1.5 in a particular dynamo sheet steel having a thickness of 0.016 in and a resistivity of 14×10^{-6} ohm-cm. In a model experiment in which he simulated a sheet of thickness 0.021 in and resistivity 14×10^{-6} ohm-cm, he found that the effects of flux waveform distortion on the actual eddy-current loss were negligible, but that, owing to this distortion, the hysteresis loss was increased sufficiently to cause an apparent extra eddy-current loss which would correspond to a loss anomaly factor of 1.18. Had the hysteresis loss of the material been greater he notes that, with the same degree of distortion, the observed anomalous loss would have been increased. The sheet simulated was both thick and of low resistivity, and these conditions are especially favourable to the promotion of flux and eddy-current waveform distortion.

In a high-quality sheet, which may have a resistivity of the order of four times that in the steel considered by Brailsford, the theoretical eddy-current loss is smaller and the change in phase of the flux density across the sheet is reduced in proportion. The amplitude of the flux density within the sheet differs from that at the surface in proportion to the square of this phase change, and the change in amplitude of flux density across the sheet will therefore be reduced more than in inverse proportion to an increase in resistivity. The hysteresis loss, depending as it does upon a power of maximum flux amplitude of 1.65 or so, will be changed in more than linear proportion to the change in the amplitude of the flux density across the sheet.

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Aspdén is with the English Electric Co., Ltd.

Thus if the resistivity is, say, doubled, the phase change of the flux density across the sheet will be halved, the change in amplitude of the flux density across the sheet will be reduced to a quarter, and it is conceivable that the apparent increase in hysteresis loss due to flux waveform distortion will be reduced to a sixth or seventh of its original value. The calculated eddy-current loss will only be halved, so that the apparent anomaly due to flux waveform distortion will be considerably reduced.

A similar argument can be applied to a decrease in thickness of the specimen, so that the degree of flux waveform distortion encountered in Brailsford's model experiments which explained an anomaly factor of 1.18 in the thick low-resistivity sheet he used is no assurance that anomaly factors as high as 3, as have been observed in thin sheet of high resistivity, can be completely explained in terms of flux waveform distortion. However, it is to be noted that the model experiment does simulate a more magnetically homogeneous sheet than may in fact exist, and flux waveform distortion leading to increased hysteresis and eddy-current loss can result from such inhomogeneity.

(2.2) Distortion in Thin High-Resistivity Sheet Steels

Although thin high-resistivity sheet steels usually have high anomaly factors, the above considerations suggest that flux waveform distortion in them is of little importance. However, it should be remembered that the actual shape of the hysteresis loop is a factor in determining this distortion, and many high-quality steels have a differential permeability over some parts of the loop which is very high, and can, in a magnetically homogeneous sheet, lead to appreciable flux waveform distortion. If the sheet is not magnetically homogeneous, it can happen that this distortion is of no consequence at all. For example, it is becoming increasingly apparent that in some sheet steels the domain configuration is such that the magnetization process comprises, in effect, the motion of a single domain boundary wall about a mean position at the centre of the sheet.^{2,4} The material surrounding this wall has a very low effective permeability and the magnetizing field is therefore more uniform across the sheet than might otherwise be expected. Furthermore, the mechanism of hysteresis is more intimately associated with the motion of the wall than effects within the surrounding material. Both these factors nullify an argument that the flux waveform is distorted across the section of the sheet. Another factor which tends to mitigate the effects of flux waveform distortion is a time-lag effect in the magnetization process.³ The effect of this time lag is to render the hysteresis loop more and more elliptical in shape as the frequency is increased. This corresponds to purely sinusoidal changes of field strength with purely sinusoidal changes of flux density, and harmonic effects do not therefore occur.

(2.3) Distortion in Thick Low-Resistivity Sheet Steels

Thick low-resistivity sheet steels usually have a low anomaly factor, and if they are magnetically homogeneous, flux-waveform-distortion effects may be expected to be important. Being relatively thick, these steels are more homogeneous magnetically than the thin steels previously considered. Furthermore, magnetic time-lag effects are known to be relatively unimportant in thick steels. This is readily explained as follows: the characteristic of an intrinsic time lag of the kind considered is that, in steels which are thick relative to the size of the ferromagnetic domains, the lag is independent of the thickness of the steel. The specific loss arising from the time lag is therefore independent of the thickness of the sheet. The corresponding classical eddy-current loss increases as the square of this thickness, and it follows that the eddy-current-anomaly component attributable to time lag decreases inversely as the square of the sheet thickness.

Flux-waveform-distortion effects are therefore likely to be of importance in thick sheet steel, and it is of interest to test such a steel with a view to resolving the observed eddy-current effects into components attributable to flux waveform distortion and other causes such as magnetic inhomogeneity and magnetic time lag.

(3) MEASUREMENT TECHNIQUE

(3.1) General Principles of Measurement

In order to carry out an experimental test which will avoid flux-waveform-distortion effects due to hysteresis phenomena, it is proposed to restrict the duration of the test so that a measurement is confined to such a small range of change of magnetization that it is reasonable to assume that the effective differential permeability is constant during the test period.

In Fig. 1, a hysteresis loop A is shown. Over the range XY the

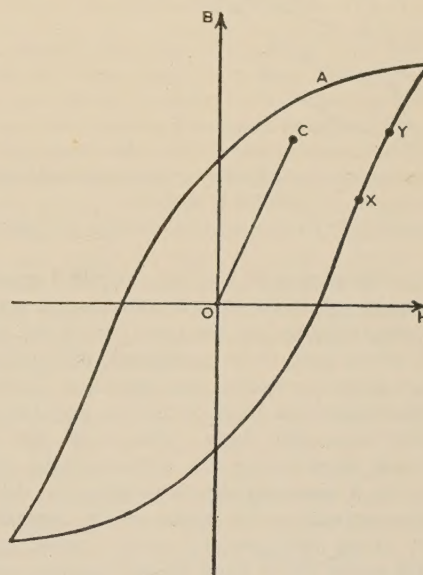


Fig. 1.—Hysteresis loop showing a small range XY over which the differential permeability is substantially constant.

differential permeability is fairly constant. It is this differential permeability which is effective in determining eddy-current action, and if this permeability is constant over the test range there should be no distortion of the eddy-current waveform, since, on a differential basis, the conditions are identical with those which would apply if the relationship between B and H had the form represented by the line OC , which is parallel to XY .

The bearing of magnetic inhomogeneity on this should be considered. Only inhomogeneity across the width of the sheet is relevant to these considerations. It is convenient to assume that the hysteresis loop A shown in Fig. 1, which is that applicable to a test specimen as a whole, is composed of two component loops D and E applicable to different layers of the sheet specimen. The loops are shown in Fig. 2. The loop A results from the loops D and E by determining the mean of the flux density B of each loop D and E for a given field strength H . It is presumed that the different loops D and E apply to equal cross-sectional areas of the test specimen. Reconsidering the range XY of the loop A, it is evident that, if the differential permeability of this loop is constant over this range, it is very likely that it is constant over the corresponding field ranges of the loops D and E. Thus there should be no eddy-current distortion over the small test range in spite of the magnetic inhomogeneity, although, of course, this inhomogeneity will affect the eddy-current loss.

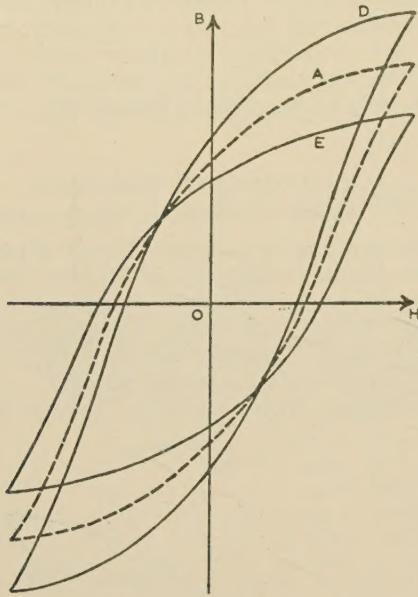


Fig. 2.—Formation of an effective hysteresis loop A from the hysteresis loops D and E.

Applicable to different layers of a non-homogeneous electrical sheet steel.

An electrical measurement can be made at a specific instant; there is no need to spread the measurement over a period of time. However, a range such as XY has been postulated because, in a magnetic test of the kind to be considered, eddy-current demagnetizing effects occur by which, at a particular instant, the field strength applicable to one layer of a sheet may be greater than that applicable to another layer. The range XY corresponds to a field range representing the demagnetizing field. Thus, for example, in a magnetically homogeneous sheet we may regard a measurement to be made at the instant when the surface layers of the sheet are at Y and a central layer is at X. The size of the range XY is fixed by the magnetic and electrical properties of the sheet and the test frequency.

If the differential permeability is constant over the range XY, then since, at the instant at which the measurement is made, all of the sheet has a differential permeability which is independent of time, no distortion can occur.

It is necessary to estimate the size of the range XY simply to ensure that the assumption based upon constant differential permeability is justified. For this purpose magnetic homogeneity can be assumed if a reasonable factor of safety is allowed. It can be shown that the demagnetizing field at the centre of the sheet is as follows:

$$H_d = \frac{2\pi b^2}{\rho 10^9} \left(\frac{\partial B}{\partial t} \right) \quad (1)$$

For a magnetization cycle in which the mean flux density within the sheet is given by

$$B = B_m \sin \omega t \quad (2)$$

eqn. (1) becomes

$$H_d = \left(\frac{2\pi b^2 \omega}{\rho 10^9} \right) B_m \cos \omega t \quad (3)$$

or putting

$$\theta = \frac{4\pi b^2 \omega \mu}{3 \rho 10^9} \quad (4)$$

$$H_d = \frac{3}{2} \theta \frac{B_m}{\mu} \cos \omega t \quad (5)$$

To assure linearity of the range XY, the smaller H_d , the better. It is zero when $\omega t = \pi/2$, but this represents a condition in which a part of the sheet has passed the loop tip and X and Y no longer define the limits of a range. The practical limit at which H is a minimum occurs when Y is at the loop tip. The mean demagnetizing field may be shown to be $\frac{3}{2} H_d$, and on the basis of eqn. (2) the flux density at the surface of the sheet can be shown to be $B_m \sin \omega t + \frac{3}{2} \mu H_d$, and that at the centre of the sheet to be $B_m \sin \omega t - \frac{3}{2} \mu H_d$. When Y is at the loop tip the value of ωt is such that $B_m \sin \omega t + \frac{3}{2} \mu H_d$ has its maximum value. From eqn. (5), this is when

$$\frac{\partial}{\partial t} (B_m \sin \omega t + B_m \theta \cos \omega t) = 0 \quad (6)$$

The maximum value is $B_m \sqrt{1 + \theta^2}$, and ωt is then approximately equal to $\pi/2 - \theta$. The position of X is determined by the corresponding value of $B_m \sin \omega t - \frac{3}{2} \mu H_d$, which, from eqn. (5), is $B_m \cos \theta - \frac{1}{2} \theta B_m \sin \theta$ or, approximately, $B_m (1 - \theta^2)$. The value of XY is then given by the difference between $B_m \sqrt{1 + \theta^2}$ and $B_m (1 - \theta^2)$, which is approximately equal to $\frac{3}{2} \theta^2 B_m$, θ being assumed to be small.

If $\theta = 0.1$ rad, the range XY is 1.5% of B_m . However, to allow for magnetic inhomogeneity of a kind which would enhance eddy-current effects, the range XY should be regarded as somewhat greater than this.

The factor θ is easy to measure, and affords a convenient theoretical expression which can readily be tested by experiment.

(3.2) Derivation of Data for Calculating Theoretical Eddy-Current Effect

To derive the theoretical value of θ it is necessary to measure the differential permeability applicable for increasing flux density just before the instant when $B = B_m$. Provided that θ is small and of the order of 0.1 rad, it is necessary that this differential permeability should be constant over a range of only 2% or so of B_m . A preliminary measurement of the value and degree of constancy of μ must therefore be made.

The well-known method of reversals formed the basis of the measurement technique for assessing μ from the static hysteresis loop. An intermediate magnetizing current is obtained by including a small series resistor in the magnetization circuit; a switch enables this resistor to be switched into and out of circuit as required. In this way a magnetizing field close to its maximum value is obtained and is changed to this value simply by short-circuiting the resistor. During this switching operation, and only then, a search coil having many turns is connected to a fluxmeter. Thus a large fluxmeter reading is obtained for a small change of field strength over a range immediately prior to B_m . This leads to a very sensitive measurement of μ . The resistance of the resistor determines the range of the test, and by varying this resistance and measuring the different values of μ , the degree of constancy of μ over the test range can be assessed. For all the results given it was found by this preliminary test that there was no noticeable variation of μ . In these tests the range was chosen to be 3% of B_m , which, as will be seen, is ample to cover the theoretical range applicable under all test conditions.

In Fig. 3 the measured values of μ applicable to a particular low-silicon steel having a mean thickness of 0.019 in and a mean resistivity of 17.5×10^{-6} ohm-cm are shown as a function of B_m . Also shown is the relationship between θ and B_m derived theoretically from these values.

It is necessary to compare this relationship with that representing a direct measure of θ .

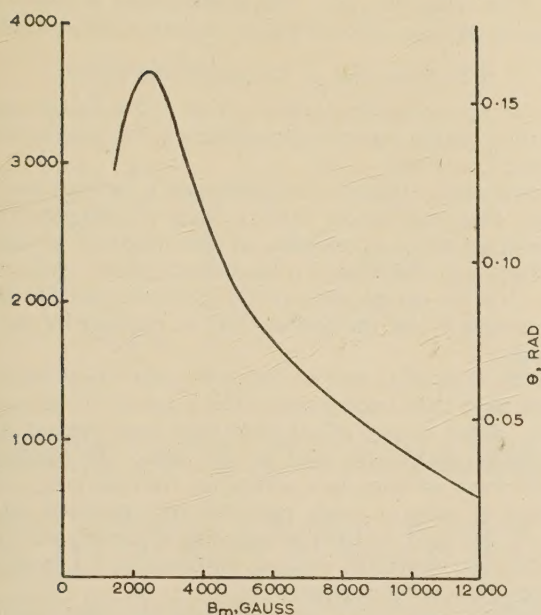


Fig. 3.—Relationship between the differential permeability μ and the maximum flux density B_m .

Applicable to a low-silicon sheet steel tested over a small range of magnetization close to B_m and corresponding values of the angle θ as given by eqn. (4) at a frequency of 50 c/s.

(3.3) Measurement of Actual Eddy-Current Effect

θ may be measured by using a magnetization cycle of which the mean flux density is varying in a purely sinusoidal manner with respect to time. The angle θ then becomes a time interval, in effect, between the instant at which the flux density at the surface of the sheet tested is a maximum and the instant at which the mean flux density within the sheet is a maximum. It is therefore necessary to measure the interval between the instants of maximum field strength at the surface of the sheet and maximum flux density within the sheet. This period was measured by advancing the phase of a signal having a 90° phase relationship with the mean flux density until a detector indicated that the field strength was a maximum when the signal was zero.

For purely sinusoidal changes of mean flux density, the e.m.f. induced in a search coil lags 90° behind the flux density and is zero when the mean flux density is a maximum. It is necessary to advance the phase of this e.m.f. by an angle θ which is such that the magnetizing current passes through a maximum when the advanced signal is zero. A uniformly magnetized closed magnetic circuit is necessary to ensure that the magnetizing current is truly proportional to the magnetizing field.

An oscilloscope was used to detect the instant at which the current maximum and e.m.f. zero occurred simultaneously. The e.m.f., which was a strong signal obtained by using a search coil having many turns, was applied directly after phase change to the X-plates of the oscilloscope. The peaks of the current waveform were chopped to produce a signal which, when amplified, was applied to the Y-plates of the oscilloscope to provide a trace of pulse form on the oscilloscope screen. A phase adjustment was then made so that the pulse moved in the X-direction until it occupied a position which was the same whether or not the e.m.f. signal supplied to the X-plates was switched on or off.

A schematic of a circuit used in these tests is shown in Fig. 4. The source of power E comprised a motor-generator set connected to a suitable filter circuit to provide an adjustable magnetizing current over a frequency range of 20–60 c/s, the output voltage having a very good sinusoidal form. The magnetizing

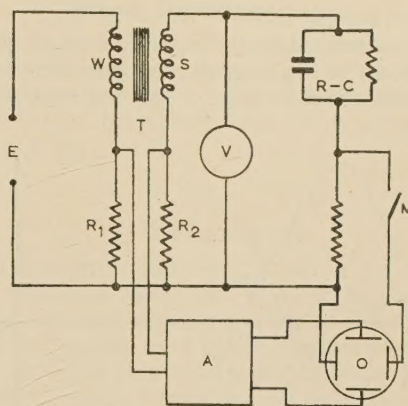


Fig. 4.—Schematic of the apparatus used to measure θ as a function of B_m and frequency.

current is passed through the magnetizing coil W , which embraces the test core T and through a resistor R_1 connected in series with the coil W . A search coil S having many turns also embraces the core T and is connected in series with a resistor R_2 . The e.m.f. induced in the search coil is measured by a voltmeter V , and an RC circuit serves to advance the phase of this e.m.f. to an adjustable extent before it is supplied through the switch M to the X-plates of the oscilloscope O . Corresponding ends of R_1 and R_2 are connected together, and a signal, derived from the other ends of the resistors, is supplied to the amplifier A . This amplifier incorporates the necessary equipment for chopping the input signal, so that output pulses corresponding to the peaks of the input signal are obtained.

It is customary in magnetic tests to have a secondary winding (search coil) having as few turns as possible, so that the effects of the secondary current are quite small. In the system shown in Fig. 4 the induced e.m.f. is required to displace the beam of a cathode-ray oscilloscope, and it must therefore be quite strong. However, whereas the possibility of amplification was considered, it was decided that the possible phase errors in an amplifier operating at frequencies less than 60 c/s introduced more complication and doubt than would be entailed if a large secondary winding were used and the trouble arising from the secondary current were compensated. The search coil S was evenly distributed on the core T . The current in this coil caused the magnetizing field not to be proportional to the current in the coil W . For this reason a compensating resistor R_2 was used. This had a resistance which bore the same ratio to the resistance of R_1 as did the number of turns of S to the number of turns of W . The potential drops in R_1 and R_2 were therefore respectively proportional to the number of ampere-turns of W and S . The difference between these potential drops constitutes the input signal to the amplifier, and this is therefore proportional to the magnetizing m.m.f. and is a measure of the magnetizing field. The resistances of R_1 and R_2 must, of course, be small in comparison with the impedances of the main windings, so as not to lead to a distortion of the magnetizing waveform.

The RC circuit was composed of standard units which were adjustable in small steps. This enabled a highly sensitive adjustment of the phase advance of the induced e.m.f. to be obtained. The overall sensitivity of the system was such that this angle could be measured to within 0.1° . The angle was obtained from a suitable locus diagram calculated on the basis of the values of the components of the RC circuit. This adjustment is made until there is no displacement of the pulse seen on the oscilloscope screen upon the switch M being opened or closed. The voltmeter V is used to measure B_m in the usual way.

(3.4) Experimental Results

In Fig. 5 measured values of θ , applicable to the steel to which the results shown in Fig. 3 apply, are presented as a function of B_m . A comparison of Figs. 3 and 5 shows that there is a

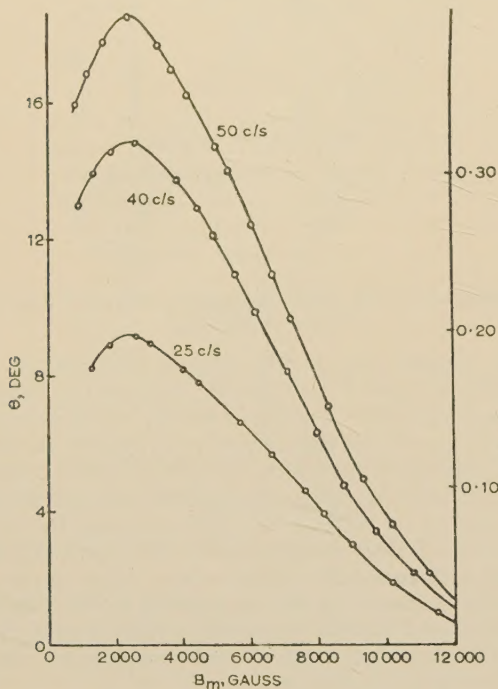


Fig. 5.—Measured values of θ as a function of B_m , with frequency as a parameter.

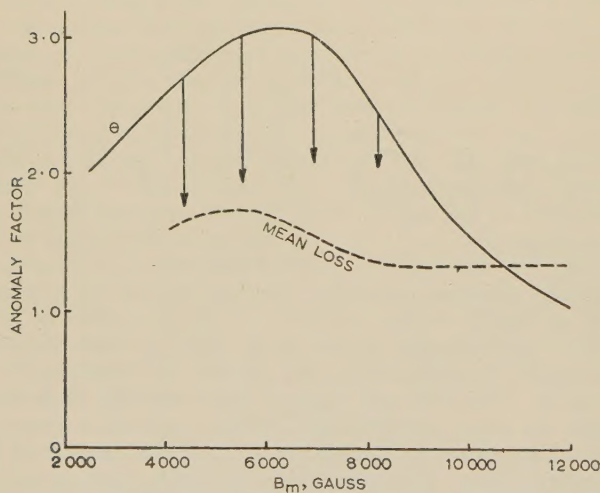


Fig. 6.—Anomaly factors applicable to the low-silicon sheet steel tested.

The full line represents the measured anomaly factors applicable over the range XY. The broken line represents the mean loss anomaly factor applicable over the whole magnetization cycle.

definite discrepancy between the results. In Fig. 6 the ratio of the measured and theoretical values of θ is shown for the 50 c/s tests.

These results are surprising, since the object of the experiments was to determine the importance of distortion effects, and by seeking to avoid such effects, it would be expected that the anomaly factor measured by these tests would be at least less, if not appreciably less, than is found by normal loss measure-

ments. This is not the case. The broken curve in Fig. 6 shows the value of the loss anomaly factor as normally measured.

(3.5) Error due to Demagnetizing Effects

There is one factor which casts a little doubt on the accuracy of the above results, but it is not sufficient to remove the element of surprise from them.

The theoretical value of θ is based upon a measurement from the static hysteresis loop. Such a loop is influenced by the demagnetizing effect of air-gaps in the magnetic circuit, and such effects are of different consequence under cyclical conditions. It is the actual permeability in a steel which is effective in determining θ , not the permeability as reduced by the effect of air-gaps.

Thus the value of μ used to derive the theoretical values of θ may well have been underestimated as a result of the action of air-gaps. Small air-gap effects must have been present, since a Lloyd-Fisher square was used for the tests. To estimate the effects of these air-gaps is a somewhat fruitless task, and the alternative of using a single picture-frame specimen was precluded by the need to have strong output signals and a pure sinusoidal flux waveform without introducing a miscellany of other possible sources of error.

The answer was found by measuring the differential permeability of the static hysteresis loop at $B = 0$. Here the differential permeability is a maximum and is usually appreciably greater than the value of μ used previously. The maximum possible effect which air-gaps could have would be to reduce an infinite actual differential permeability to the observed maximum differential permeability. The same air-gap effect upon the differential permeability μ_a is to reduce it to $\mu_a/(1 + \mu_a/\mu_m)$, where μ_m is the maximum differential permeability at the point $B = 0$. When the maximum possible air-gap effect is accounted for in this way, the full curve shown in Fig. 6 is displaced to the positions marked by the arrow-heads. Even this does not bring the anomaly factor measured near B_m to a value less than the loss anomaly factor. Air-gap effects of this magnitude cannot, in fact, occur, but the analysis does show their nature and how important they can be in these particular tests.

The general result obtained is sufficient to indicate that there is no evidence that distortion plays an important part in the eddy-current anomaly. The anomaly must therefore be attributed to other causes, such as magnetic inhomogeneity.

(4) ANALYSIS OF RESULTS

The actual magnetic domain structure of the sheet steel tested has been regarded as accounting for a magnetic inhomogeneity sufficient to explain the anomaly factors shown in Fig. 6. This cannot be verified directly because, even though it may be possible to observe the actual domain structure in the steel, it is impossible to assess its mean effect, especially under dynamic magnetization conditions.

However, various hypothetical domain structures may be considered, and the one selected which best satisfies the actual results obtained. This is not proof that such a domain structure exists, but the fact that the measured anomaly factors can be explained in terms of the domain pattern affords an acceptable basis for the explanation of the eddy-current anomaly.

The particular results presented seem to be best explained by a hypothetical domain configuration of the form shown in Fig. 7. Here, there are three parallel domain boundaries XX, YY, ZZ. Parts of the sheet between adjacent boundaries and between a boundary and an adjacent surface of the sheet are assumed to be magnetized to saturation, in the directions indicated by the arrows. The thickness of the sheet is $2b$, and the boundaries

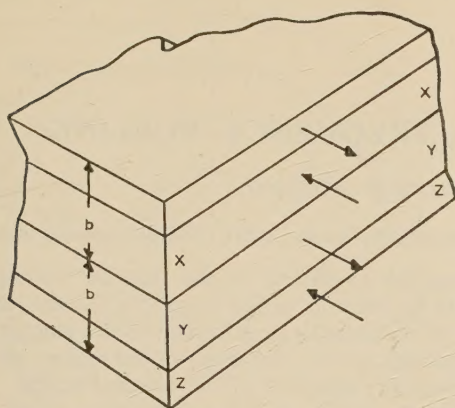


Fig. 7.—Hypothetical domain configuration used as a basis for an explanation of the anomaly effects observed in the steel tested.

XX and ZZ are presumed to be distant $2b/3$ from the central boundary YY. The resultant flux density of the system is therefore zero.

A resultant flux is established by displacing the boundaries by a magnetic field. If all three boundaries are displaced at an equal rate it may be shown that the anomaly factor is 1.15. If the boundaries XX and ZZ are displaced at an equal rate whilst the boundary YY remains still, it can be shown that the anomaly factor is 1.0, and if the boundary YY is presumed to move, whilst the boundaries XX and ZZ remain still, it can be shown that the anomaly factor is 3.0. These anomaly factors change according to the displacement of the moving boundaries from their neutral positions.

It is likely that, for rapid rates of flux change such as occur when the mean flux density passes through zero during a magnetization cycle, the eddy-current demagnetizing effects will act to screen the boundary YY. The result is that a relatively low anomaly factor will be observed. For low rates of flux change, as occur over the range of test used in the above experiments, this screening effect will be small, and the boundary YY can be expected to move with the others. The mean flux density will then have the value B_m , and the boundaries XX and ZZ will be displaced to new positions, one having moved towards the centre of the sheet and the other being close to the surface and very likely more difficult to displace. Almost all the flux change therefore occurs near the centre of the sheet, and the anomaly factor may approach the value 3.0.

The mean loss anomaly factor over a complete cycle of magnetization can be expected to be appreciably less than 3.0.

In support of the above hypothetical explanation it is to be noted that the anomaly factor has been measured at the instant when the mean flux density is zero during a cycle of magnetization, and has been found to have a value of 1.04 in the steel tested above.² This is consistent with the hypothetical domain pattern shown in Fig. 7.

(5) CONCLUSIONS

A method has been described by which the eddy-current anomaly in electrical sheet steel can be measured without involving flux-waveform-distortion effects. It has been shown that, instead of leading to a closer agreement with classical eddy-current theory as is to be expected from an elimination of distortion effects, higher anomaly factors were encountered. From this, it was apparent that the anomaly factor varied during a cycle of magnetization. The magnitudes of the anomaly factors measured were sufficient to show that distortion effects are relatively unimportant as a cause for the eddy-current anomaly, even though the particular steel tested was thick and had a low resistivity. It has been found possible to explain the observed variation of the eddy-current anomaly in terms of a hypothetical magnetic inhomogeneity arising from the domain configuration.

(6) ACKNOWLEDGMENTS

The author acknowledges a sincere debt of gratitude for the awards of a Whitworth Senior Scholarship and The Institution I.M.E.A. Research Scholarship which enabled him to undertake this work. He is also grateful for the facilities and kind assistance afforded him at the Cambridge University Engineering Laboratories, and particularly the helpful encouragement shown him by Mr. G. E. Middleton.

(7) REFERENCES

- (1) BRAILSFORD, F.: 'Investigation of the Eddy-current Anomaly in Electrical Sheet Steels', *Journal I.E.E.*, 1948, **95**, Part II, p. 38.
- (2) ASPDEN, H.: 'The Eddy-Current Anomaly in Electrical Sheet Steel', *Proceedings I.E.E.*, Monograph No. 164 M, January, 1956 (**103 C**, p. 272).
- (3) ASPDEN, H.: 'Magnetic Time-Lag Effects in Solid Steel Cores', *ibid.*, Monograph No. 165 M, January, 1956 (**103 C**, p. 279).
- (4) BOZORTH, R. M.: 'Ferromagnetic Domains', *Physica*, 1949, **15**, p. 207.

THE MEASUREMENT OF MAGNETRON FREQUENCY PULLING

By J. R. G. TWISLETON, B.Sc., Associate Member.

(The paper was first received 2nd March, and in revised form 24th April, 1956. It was published as an INSTITUTION MONOGRAPH in July, 1956.)

SUMMARY

The frequency pulling of an oscillating magnetron is measured by inserting a reactor in the output feeder, which produces a reflection of constant magnitude but variable in phase; the frequency pulling is usually estimated on the assumption of constant v.s.w.r. in the feeder. In practice, when multiple reflected waves exist in the feeder the v.s.w.r. is not constant, but it is found that the amplitude/phase variation of the resultant reflected wave is such that the locus of the voltage reflection coefficient is a circle in the Smith chart. An idealized equivalent circuit for the magnetron is assumed and the theoretical frequency pulling is estimated. The performance of a typical waveguide system used in magnetron testing is examined with regard to the possible variation in measured frequency pulling.

LIST OF SYMBOLS

- f_p = Frequency pulling of magnetron.
 Q_x = External Q-factor of magnetron.
 f = Frequency of oscillation.
 S = Voltage standing-wave ratio due to waveguide mismatcher = $\frac{1 - |r|}{1 + |r|}$, where r is the corresponding voltage reflection coefficient.
 $\frac{Q_x}{Q_0} + j2Q_x \frac{\delta f}{f_0}$ = Admittance of magnetron resonant circuit.
 Q_0 = Internal Q-factor of magnetron.
 $\delta f/f_0$ = Fractional frequency deviation expressed in terms of the resonant frequency, f_0 .
 $G + jB$ = Admittance presented by waveguide load.
 R = Radius of frequency-pulling circle.
 $Ce^{j\phi}$ = Radius vector to centre of frequency-pulling circle.
 $\Delta f_p/f_{p0}$ = Fractional variation in frequency pulling.
 λ_g = Wavelength in waveguide.
 λ_0 = Wavelength in free space.
 $\delta \lambda_g$ = Change in guide wavelength corresponding to change in frequency δf .
 n = Distance of mismatcher from magnetron in number of guide wavelengths.

(1) INTRODUCTION

The measurement of frequency pulling is an important feature in magnetron oscillation tests, and experiments with mismatched waveguide loads yield a class of data which is of interest to the magnetron designer and user.¹ In manufacture, the measurement of frequency pulling is carried out as a routine test.

While the frequency of an oscillating magnetron is determined almost entirely by the resonant frequency of its anode resonator, the reactive impedance presented by the external load produces small changes in frequency which are inversely proportional to the external Q-factor. This situation arises when the output feeder or waveguide is mismatched, and the change in frequency depends upon the magnitude and phase of the reflected wave presented to the magnetron. If in terms of unit transmitted

voltage wave the amplitude of the reflected voltage wave is constant, but its phase is varied from 0 to 2π rad, the reactive component of the impedance presented to the magnetron varies in a cyclic manner and the frequency of the magnetron oscillates about the frequency obtained with a matched feeder.

The frequency pulling of the magnetron, i.e. the difference between the maximum and minimum frequencies observed, may be shown to be²

$$f_p = \frac{f}{Q_x} \frac{2|r|}{1 - |r|^2} = \frac{f}{Q_x} \left(\frac{1}{S} - S \right) \quad \dots \quad (1)$$

The frequency pulling figure (f.p.f.) is defined as the total frequency change obtained when $|r| = 0.2$ (or $S = 0.666$).

Therefore
$$\text{f.p.f.} = 0.417 \frac{f}{Q_x} \quad \dots \quad (2)$$

It is assumed in the derivation of eqn. (1) that the voltage standing-wave ratio in the output feeder is constant. This implies that the locus of the voltage reflection coefficient in the Smith chart, which is traced out when the phase of the reflected wave is varied from 0 to 2π rad, is a circle centred at the point $1 + j0$. The change in phase of the reflected wave with frequency, due to the change in electrical length of the feeder between the source of the reflected wave and the magnetron, is also neglected, but as is shown in Section 7.1, the correction to be applied to eqn. (1) is small when the feeder is short.

In practice, the feeder and the termination used for oscillation tests consists of a short waveguide run comprising a mismatcher terminated by a water load. The mismatcher is arranged to produce a reflected wave which is variable in magnitude and phase. In the measurement of frequency pulling only the phase is varied, as previously described. However, owing to imperfections in the mismatcher, or as a result of reflected waves which arise in other parts of the waveguide system, the magnitude of the resultant reflected wave, in terms of unit incident wave, may not be independent of phase. In this case the value of $|r|$ to be used in eqn. (1) becomes uncertain. As an illustration, consider the behaviour of the waveguide system with two types of mismatcher which are used for magnetron frequency-pulling tests, namely a travelling probe and a fixed probe preceded by a phase-shifter.³

(1.1) Behaviour of the Waveguide System

The travelling-probe mismatcher consists of a length of waveguide having a dielectric probe projecting into the waveguide. The magnitude of the wave reflected by the mismatcher depends upon the penetration of the probe, and its phase is varied directly by traversing the probe along a slot cut in the wall of the waveguide. Subsidiary reflected waves may be produced at the ends of this slot, or by the couplings at each end of the section of waveguide, so that the reflected wave presented to the input of the mismatcher is the resultant of several component waves. The phase differences between the subsidiary reflected waves and the main reflection due to the probe are determined by the position of the probe in the slot, and hence the magnitude of the resultant wave is found to depend upon the linear position of the probe.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Twisleton is with the British Thomson-Houston Co., Ltd.

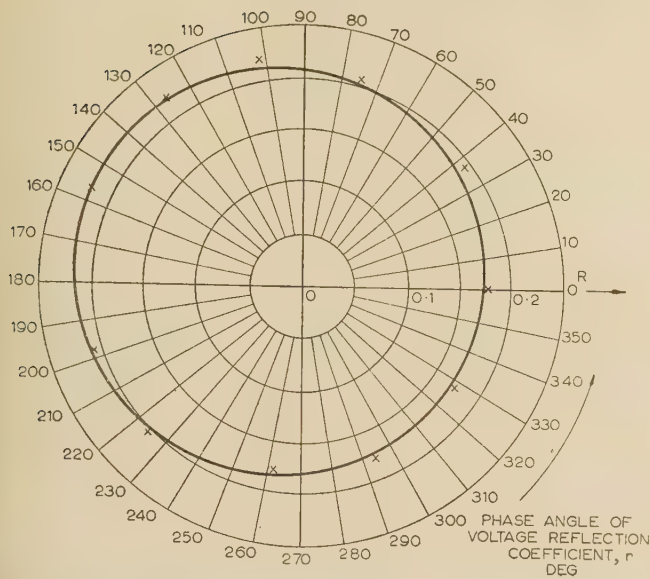


Fig. 1.—Circle diagram of voltage reflection coefficient for X-band travelling-probe mismatch in $1 \text{ in} \times \frac{1}{2} \text{ in}$ rectangular waveguide.

The performance of an X-band travelling-probe mismatch when terminated with a matched load is shown in Fig. 1. The plot shows the locus of the complex voltage reflection coefficient resulting from a linear traverse of the probe; this locus is a circle known as the pulling circle, and it will be seen that the centre is displaced from the origin of the diagram.

The size and position of this circle will also depend upon the reflection produced by the water load or by the reflections arising in the sections of waveguide preceding it, so that, in general, the pulling circle for the complete waveguide system will be off-centre in the Smith chart.

The fixed-probe mismatch produces a reflected wave in the same manner as the travelling-probe type, except that the phase of the reflected wave is varied by means of a phase-shifter preceding the fixed probe. A well-known type of phase-shifter³ consists of a length of waveguide having an internal wedge of insulating material which can be displaced transversely across the guide. Reflected waves are produced in the dielectric-filled section of the guide at the changes in the cross-section of the wedge, and these interfere when the wedge is moved across the guide, owing to the change in guide wavelength.

In practice, the fixed-probe mismatch described is superior to the travelling probe, since the reflected wave due to the phase-shifter can be made fairly small, with the result that the magnitude of the resultant reflected wave is almost independent of the setting of the phase-shifter. However, it is still necessary to consider reflections which arise in the length of guide between the mismatch and the magnetron. Generally, the pulling circle produced by the waveguide system will be off-centre in the Smith chart.

The frequency pulling of the magnetron depends upon the position of the pulling circle in the Smith chart, observed at a plane in the feeder corresponding to the equivalent circuit of the magnetron. The position of such a plane is usually unknown, and it varies with each particular magnetron.

To obtain an estimate of the possible error in the measured frequency pulling we consider the theoretical frequency pulling which is obtained when the pulling circle is off-centre in the Smith chart, and the variation in frequency pulling which results from a rotation of the centre of the pulling circle; this is equivalent to a shift in the terminal plane of the magnetron.

The performance of a typical waveguide system used for magnetron tests is examined, and means for minimizing unwanted mismatches are discussed.

(2) THEORETICAL FREQUENCY PULLING

The magnetron is represented by a parallel resonant circuit, shown in Fig. 2, having shunt impedances which are related to

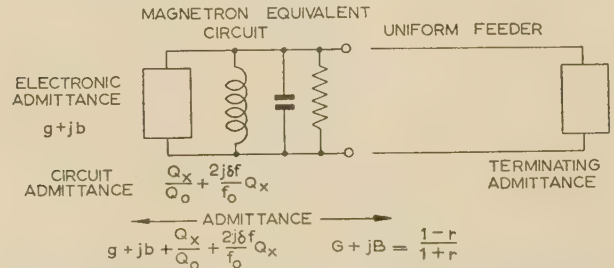


Fig. 2.—Equivalent circuit for magnetron and waveguide load.

the Q -factors, and an additional non-linear admittance $g + jb$, which is due to the electron stream;⁴ this circuit is valid at frequencies near to the resonant frequency, provided that other resonances are far removed. The admittance of this equivalent circuit, expressed in terms of the Q -factors and normalized to the characteristic admittance of the feeder, is

$$y = \frac{Q_x}{Q_o} + j2Q_x \frac{\delta f}{f_o} + g + jb \quad (3)$$

If the feeder is mismatched so that the admittance presented to the magnetron is $G + jB$, the frequency shift δf is given by

$$b + 2 \frac{\delta f}{f_o} Q_x = -B \quad (4)$$

Provided that b is constant,* the frequency pulling which occurs when the phase of the reflected wave is varied through 2π rad is

$$f_p = \frac{f_o}{2Q_x} (B_{max} - B_{min}) \quad (5)$$

Since the relationship between r and y is bilinear, it follows that the locus of y is also a circle and that the total change in susceptance is equal to the diameter of the circle in the admittance plane. Therefore

$$f_p = \frac{f_o}{Q_x} A \quad (6)$$

where A is the radius of the admittance circle.

Since the locus of r is a circle in the complex plane, we may put

$$r = C e^{j\phi} + R e^{j\theta} \quad (7)$$

where $C e^{j\phi}$ is the radius vector to the centre of the circle $R e^{j\theta}$ shown in Fig. 3. The phase angles ϕ and θ are measured at the

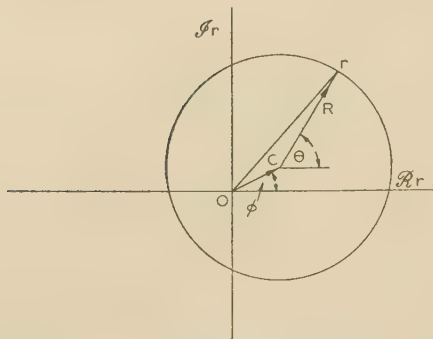
terminals of the equivalent circuit. Since $y = \frac{1-r}{1+r}$,

$$\frac{1-y}{1+y} - C e^{j\phi} = R e^{j\theta} \quad (8)$$

Eqn. (8) yields the equation of the admittance circle by eliminating θ as follows:

$$\left(\frac{1-y}{1+y} - C e^{j\phi} \right) \left(\frac{1-\bar{y}}{1+\bar{y}} - C e^{-j\phi} \right) = R^2$$

* It is found experimentally that b is a function of g , so that b is only approximately constant. We omit this consideration here.

Fig. 3.—Circle diagram for voltage reflection coefficient r .

where \bar{y} is the conjugate of y . When this equation is multiplied out and equated to zero, an equation of the type given in Section 7.2 is obtained [eqn. (29)] having terms in $y\bar{y}$, y and \bar{y} , plus a constant.

The factor of \bar{y} is

$$V = -\frac{(C^2 - R^2 - 1 + 2jC \sin \phi)}{1 + C^2 - R^2 + 2C \sin \phi} \quad (9)$$

and the constant term is

$$V\bar{V} - A^2 = \frac{1 + C^2 - R^2 - 2C \cos \phi}{1 + C^2 - R^2 + 2C \cos \phi} \quad (10)$$

Eliminating $V\bar{V}$ from the latter equation we find that the radius of the admittance circle is

$$A = \frac{2R}{1 + C^2 - R^2 + 2C \cos \phi} \quad (11)$$

The frequency pulling, from eqn. (7), is

$$f_p = \frac{f_0}{Q_x} \frac{2R}{1 + C^2 - R^2 + 2C \cos \phi} \quad (12)$$

Eqn. (12) shows that the frequency pulling attains maxima and minima when $\cos \phi = -1$ and $\cos \phi = 1$, respectively, giving

$$(f_p)_{\max} = \frac{f_0}{Q_x} \frac{2R}{1 + C^2 - R^2 - 2C} \quad (13)$$

and

$$(f_p)_{\min} = \frac{f_0}{Q_x} \frac{2R}{1 + C^2 - R^2 + 2C}$$

In both cases the centre of the circle in the Smith chart lies on the real axis, towards the point $\pm j\infty$ when maximum frequency pulling occurs, and towards the point $j0$ when minimum frequency pulling occurs.

If f_{p0} is the frequency pulling obtained when the circle is central in the Smith chart, then

$$f_{p0} = \frac{f_0}{Q_x} \frac{2R}{1 - R^2} \quad (14)$$

This equation corresponds to eqn. (1). Therefore, the ratio f_p/f_{p0} (f_{p0} being the value of frequency pulling obtained in the ideal case) is

$$\frac{f_p}{f_{p0}} = \frac{1}{1 + \frac{C(C + 2 \cos \theta)}{1 - R^2}} \quad (15)$$

and the fractional error in f_p due to the circle being off-centre by an amount C is approximately

$$\frac{\Delta f_p}{f_{p0}} = \frac{C(C + 2 \cos \theta)}{1 - R^2} \quad (16)$$

provided that $\Delta f_p/f_{p0}$ is small.

Neglecting the term $-C^2$ in the numerator, the maximum and minimum values of $\Delta f_p/f_{p0}$ are

$$\frac{\Delta f_p}{f_{p0}} = \pm \frac{2C}{1 - R^2} \quad (17)$$

The error in frequency pulling may be zero for a particular position of the r circle in the Smith chart. If ϕ' is the value of ϕ such that this error is zero, equating eqns. (12) and (14) gives

$$C(C + 2 \cos \phi') = 0$$

Therefore $\cos \phi' = -C/2$.

The latter equation gives $\phi' = \pm \left(\frac{\pi}{2} + \frac{C}{2}\right)$ provided that

C is small. Thus, in this case the circle may lie in the Smith chart on either side of the real axis and slightly towards the point $\pm j\infty$.

(2.1) Calibration of the Waveguide System

The mismatch probe is inserted to produce a voltage reflection coefficient of the desired magnitude, and the phase is varied through all possible values by means of the phase changer. If S_1 and S_2 are the maximum and minimum voltage standing-wave ratios, both less than unity,

$$\frac{1 - S_1}{1 + S_1} = R + C$$

and

$$\frac{1 - S_2}{1 + S_2} = R - C$$

provided that $R > C$. This is so if the total change in the phase of the reflected wave is greater than π rad, as the v.s.w.r. assumes all its values. (If the total phase change is zero, $R < C$; and if it is π rad, $R = C$.)

Then

$$R = \frac{1 - S_1 S_2}{1 + S_1 + S_2 + S_1 S_2} \quad (18)$$

and

$$C = \frac{S_1 - S_2}{1 + S_1 + S_2 + S_1 S_2}$$

Thus, knowing S_1 and S_2 , the maximum error which is possible in a measurement of frequency pulling, due to unwanted mismatches in the waveguide system, may be calculated from eqns. (13) or (17).

(3) PERFORMANCE OF A TYPICAL WAVEGUIDE SYSTEM USED FOR MAGNETRON TESTING

The v.s.w.r. produced by a waveguide system comprising a polythene-wedge water load, a mismatch consisting of a Distrene rod penetrating the narrow face of the guide, a dielectric-slab phase-shifter and an H-plane bend, was measured using a low-power test bench. These components were assembled in the above order, and thin sheets of mica were used between the waveguide flanges at each end of the phase shifter to permit a higher pressure when testing magnetrons. The internal dimensions of the waveguide were 2.37 in by 1.12 in.

The variation in v.s.w.r. was from 0.97 to 0.90 without the Distrene rod inserted, and from 0.68 to 0.62 with the rod inserted, over the total range of phase-shift. Thus, from

eqn. (18), $R = 0.213$ and $C = 0.022$, so that according to eqn. (17) the maximum error in the measurement of frequency pulling, due to unwanted mismatches in the waveguide system, would be $\pm 4.6\%$.

An attempt was made to reduce this error, and the v.s.w.r. due to each component was measured separately, when terminated with a matched load.

The residual reflection in each component, excluding the phase-shifter, was matched out using standard techniques, e.g. the reflection produced by the sheets of mica was compensated by an inductive iris in the plane of the sheet. The final v.s.w.r. produced by each section of waveguide was better than 0.99.

The v.s.w.r. produced by the complete waveguide system was then 0.952–0.938 without the Distrene rod inserted, and 0.638–0.645 with the rod inserted over the total range of phase-shift. The circles of voltage reflection coefficient which were obtained

the variation constituting the maximum error in the measurement due to the waveguide system. It depends upon the magnitude of subsidiary reflections relative to the reflection due to the mismatch, and also the variation in v.s.w.r. produced by the mismatch alone. When the unwanted reflections are minimized, the variation in frequency pulling depends only upon the variation in v.s.w.r. produced by the mismatch. With the type of phase shifter described the error due to the waveguide system can, with care, be made as small as that involved in the measurement of frequency differences.

It should be mentioned, however, that the characteristics of most types of water load are not sufficiently stable for the calibrations to hold exactly over long periods, and hence the average degree of accuracy attainable in the measurement is likely to be less than that indicated in the paper.

(5) ACKNOWLEDGMENTS

This investigation was carried out under a contract placed by the Department of Physical Research, Admiralty, and thanks are given to the Admiralty for permission to publish the paper. The author is also indebted to his colleagues in the B.T.-H. Research Laboratory for encouragement and assistance in this work, particularly Mr. T. H. B. Baker.

(6) REFERENCES

- (1) COLLINS, G. B.: 'Microwave Magnetrons' (McGraw-Hill, New York, 1948), Chapter 7.
- (2) LYTHALL, B. W.: 'Frequency Instability of Pulsed Transmitters with Long Waveguides', *Journal I.E.E.*, 1946, 93, Part IIIA, p. 1081.
- (3) HALFORD, G. J.: 'A Wide-Band Waveguide Phase-Shifter', *Proceedings I.E.E.*, Paper No. 1466, May, 1953 (100, Part III, p. 117).
- (4) SLATER, J. C.: 'Microwave Electronics' (Van Nostrand, New York, 1950), Section 9.2.

(7) APPENDICES

(7.1) The Frequency Pulling with a Short Lossless Feeder

In the derivation of eqn. (1) it is assumed that the source of reflection in the output feeder is sufficiently close to the magnetron for the change in phase of the reflected wave with frequency to be neglected. However, in a discussion of the influence of other properties of the feeder upon the measured frequency pulling it is desirable to estimate also the effect of the finite length of the feeder.

It is assumed here, as in Section 2, that the frequency of the magnetron is determined only by the susceptance at a plane in the output feeder near the magnetron, which may be regarded as the location of the magnetron equivalent circuit.

The behaviour of the magnetron can then be estimated when the reflected wave is produced at the far end of the feeder by calculating the susceptance at the near end. Because the derivation is approximate, the case of constant v.s.w.r. in the feeder only is considered.

If the complex voltage reflection coefficient at the far end of the feeder is $Re^{j\theta}$, the corresponding voltage reflection coefficient at the near end (the plane of the magnetron equivalent circuit) is

$$r = Re^{j(\theta - \frac{4\pi l}{\lambda_g})} \quad (19)$$

where l is the length of the feeder.

If the frequency is changed slightly such that the new guide wavelength becomes $\lambda_g + \delta\lambda_g$, then

$$r = Re^{j[\theta - 4\pi n(1 - \frac{\delta\lambda_g}{\lambda_g})]} \quad (20)$$

where $n = l/\lambda_g$.

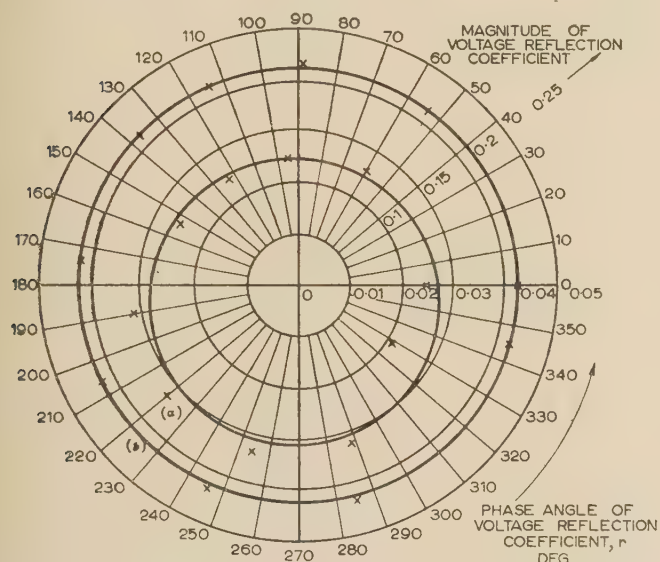


Fig. 4.—Circle diagram of voltage reflection coefficient for waveguide system.

- (a) Mismatch slug withdrawn; radial scale 0–0.05.
(b) Mismatch slug inserted; radial scale 0–0.25.

are shown in Fig. 4. The radial displacement of the circle obtained with the Distrene rod inserted is seen to be negligible. The inner circle, obtained with the rod withdrawn, showed that the residual mismatch due to the phase-shifter and bend was very small, since this circle encloses the origin.

With this waveguide system, the frequency pulling of a pulsed magnetron was measured using a spectrometer display with 1 Mc/s frequency markers. The magnetron peak current was chosen such that the frequency variation with current was zero. The frequency pulling with the Distrene rod inserted was 6.8 Mc/s, and with the rod withdrawn was 0.8 Mc/s.

Assuming the low-power calibration to hold, the ratio of pulling figures should be, from eqn. (1),

$$\frac{0.216}{0.027(1 - 0.046)} = 8.24$$

The ratio agrees well with the ratio of the measured frequency pulling, i.e. $6.8/0.8 = 8.5$.

(4) CONCLUSIONS

With the type of waveguide system described a variation of $\pm 5\%$ in the measured value of frequency pulling is probable,

The change in guide wavelength, $\delta\lambda_g$, and the change in frequency, δf , are related as follows:

$$\frac{\delta\lambda_g}{\lambda_g} = -\frac{\lambda_g^2}{\lambda_0^2} \frac{\delta f}{f}$$

so that eqn. (20) becomes

$$r = R e^{j(\psi - 4\pi n \frac{\lambda_g^2 \delta f}{\lambda_0^2 f})} \quad (21)$$

where ψ , which is the phase of the reflected wave observed at the near end, is substituted for $\theta - 4\pi n$.

The susceptance at the plane of the magnetron equivalent circuit can be found using eqn. (8), and the equation relating the frequency shift of the magnetron δf and the phase of the reflected wave is

$$\frac{\delta f}{f} = \frac{1}{Q_x} \frac{R(\sin \psi \cos \chi - \sin \chi \cos \psi)}{1 + R^2 + 2R(\cos \psi \cos \chi + \sin \chi \sin \psi)} \quad (22)$$

where

$$\chi = 4\pi n \frac{\lambda_g^2 \delta f}{\lambda_0^2 f} = k \frac{\delta f}{f}$$

If χ is much less than unity, which is the case when the feeder is short, then putting $\cos \chi = 1$, and $\sin \chi = \chi$, eqn. (22) becomes explicitly

$$\frac{\delta f}{f} = \frac{1}{Q_x} \frac{R \sin \psi}{1 + R^2 + \left(2 - \frac{k}{Q_x}\right) R \cos \psi} \quad (23)$$

The frequency shift, δf , has a maximum value when

$$\cos \psi = -\frac{\left(2 - \frac{k}{Q_x}\right) R}{1 + R^2}$$

so that, substituting this value for $\cos \psi$ in eqn. (23), the frequency pulling is found to be

$$f_p = \frac{f}{Q_x} \frac{2R}{\sqrt{(1 + R^2)^2 - \left(2 - \frac{k}{Q_x}\right)^2 R^2}} \quad (24)$$

This formula may be written, approximately,

$$f_p = \frac{f}{Q_x} \frac{2R}{1 - R^2} \left[1 - \frac{kR^2}{Q_x(1 - R^2)^2}\right] \quad (25)$$

which can obviously be expressed in terms of the frequency pulling, now designated f'_p , given by eqn. (1). Thus

$$f_p = f'_p \left(1 - n\pi \frac{\lambda_g^2}{\lambda_0^2} \frac{f_p'^2 Q_x}{f^2}\right) \quad (26)$$

In the practical case examined in Section 3 the length of waveguide between the mismatch and the magnetron was approximately four guide wavelengths, so that the correction term in eqn. (26) is 0.019, since $\lambda_g/\lambda_0 = 12.55$, $f'_p/f = 0.00184$, and $Q_x = 244$.

Thus, in this instance the correction which has to be applied to the frequency pulling, as given by eqn. (1), is only -2%. This would be less than the experimental error involved in the measurement.

(7.2) The Equation of a Circle in the Complex Plane

The equation of a circle in the x, y plane is

$$x^2 + y^2 + g^2 + f^2 - A^2 - 2gx - 2fy = 0 \quad (27)$$

where g and f are the x and y co-ordinates of the centre of the circle and A is the radius of the circle.

Let $z = x + jy$ be a point in the complex plane and $V = g + jf$ be the centre of the circle in the complex plane.

$$\left. \begin{aligned} \text{Then } x &= \frac{1}{2}(z + \bar{z}) \text{ and } g = \frac{1}{2}(V + \bar{V}) \\ y &= \frac{1}{2j}(z - \bar{z}), f = \frac{1}{2j}(V - \bar{V}) \\ x^2 + y^2 &= z\bar{z}, g^2 + f^2 = V\bar{V} \end{aligned} \right\} \quad (28)$$

where \bar{z} is the conjugate of z and \bar{V} is the conjugate of V .

Substituting for x, y, g , and f from eqn. (28), in eqn. (27) we find that the equation of the circle in the complex plane is

$$z\bar{z} - z\bar{V} - \bar{z}V + V\bar{V} - A^2 = 0 \quad (29)$$

A METHOD FOR THE APPROXIMATE DETERMINATION OF THE IMPULSE RESPONSE OF A NUMBER OF IDENTICAL CIRCUITS IN CASCADE

By K. F. SANDER, M.A., Ph.D., Associate Member.

(The paper was first received 18th February, and in revised form 23rd April, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

The method of steepest descents is applied to evaluate the impulse response of a number of identical buffered circuits in cascade. It is shown that exact results may be obtained by analogue computing. Approximate analytical expressions are obtained for various values of time, including the period of build-up. These are tested against two networks for which exact evaluation is possible. A cruder approximation valid near the build-up time is applied to predict suitable parameters for video-frequency compensating networks. These again are checked with known results.

LIST OF PRINCIPAL SYMBOLS

- $p = \xi + j\omega$, variable of Laplace transform.
 $z = x + jy$, reduced variable of the form p/ω_0 .
 $\left. \begin{matrix} \xi \\ w \end{matrix} \right\} =$ General complex variables.
 t = Time.
 τ = Reduced time of form $\omega_0 t$.
 $G(p)$ = Gain function of an amplifier.
 p_s, q_s = Zeros and poles of G .
 $X(p) = \text{Log } G(p) = \alpha + j\beta$.
 $A_n(t)$ = Impulse response of n stages.

$$F = \frac{pt}{n} + X = \frac{z\tau}{n} + X = \phi + j\psi$$

 F_c = Value of F at a col.
 $\sigma = F_c - F$.
 $\lambda = \tau/n$.
 $\alpha_1 = X_c''/2$ (dashes denote differentiation).
 $\alpha_2 = X_c'''/6$.
 $\mu = n^{2/3} \alpha_1^2 / (3\alpha_2)^{4/3}$.
 $J_\nu(x)$ = Bessel function.
 $\text{Ai}(x)$ = Airy function.
 $\left. \begin{matrix} h_1(x) \\ h_2(x) \end{matrix} \right\} =$ Modified Hankel function of order one-third.
 $\Gamma(x)$ = Gamma function.

Various symbols, including some of those above, are used as constants or parameters in places where no confusion may be caused, and are defined in such places.

(1) INTRODUCTION

The problem of determining the transient response of a network is usually simple in theory but difficult in practice, in all except the simplest cases. Numerical methods of evaluating the integrals obtained are available, although these are liable to be tedious when large phase shifts are involved. Several examples of numerical methods are given in Reference 1. The development of general analytical expressions is necessary to establish the more detailed relationships between gain/frequency curves

and transient response. The use, for this purpose, of the physically unrealizable example of a square pass band associated with linear phase shift is unsatisfactory. A cascade of identical circuits is a comparatively simple form of network for which analytical results may be derived, and, apart from the general importance of the results, it is also a case of practical importance, particularly in connection with wide-band amplifiers.

Work on multi-stage circuits has been comparatively small. That most similar to the present paper is due to Carson,² who obtained approximate expressions valid for very long filters occurring in long-distance transmission, covering some hundreds of loaded sections. Eaglesfield³ obtained expressions for cascaded single-tuned circuits. Both of these results were obtained in the terms of envelopes of a carrier waveform.

In the present paper the technique of transformation of contour integrals by the method of steepest descents is used to obtain approximate expressions for the impulse response when the circuits have a more general form. Particular cases are examined in detail to obtain estimates of error.

It may be that the use of modern electronic computing machines affords the best method for numerical evaluation of the transient response of large numbers of stages. The differential equations are of simple form, although simultaneous, and repeated solutions with the output derived from the preceding stage of computation would be very simple to programme. No work appears to have been done on these lines.

(2) THE IMPULSE RESPONSE OF CASCADED AMPLIFIER STAGES

(2.1) Formulation of the Problem

The simplest problem of many stages of a given circuit is that occurring in a multi-stage amplifier without feedback. Here the voltage transfer ratio between the inputs of consecutive stages is independent of the number of stages, in contrast to the problem of a filter with a finite number of stages. The analysis developed below produces expressions for the response of a multi-stage amplifier of this character in the form of an asymptotic expansion valid for a large number of stages. In certain practical cases of wide-band amplifiers as many as twenty stages are used, and a suitable expansion can be expected to give a reasonably accurate answer.

If a circuit is such that the ratio of the Laplace transforms of the output and input voltages is given by $G(p)$ the response to a unit impulse is given by the inversion integral

$$A(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p) e^{pt} dp \quad \dots \quad (1)$$

where all singularities of $G(p)$ lie to the left of the contour. For a real amplifying stage, stray capacitances will ensure that response to a sine wave decreases as frequency increases, and in fact $G(p) \simeq G_0/p^N$ for sufficiently large values of p . For a

Correspondence on Monographs is invited for consideration with a view to publication.
 Dr. Sander is in the Department of Engineering, University of Cambridge.

finite number of elements it will be possible to write $G(p)$ in the form

$$G = G_0 \frac{\prod_{r=1}^{N_1} (p - p_r)}{\prod_{s=1}^{N_2} (p - q_s)} \simeq \frac{G_0}{p^{N_2}} \quad |p| \gg 1$$

where $N = N_2 - N_1$.

p_r and q_s are the zeros and poles of $G(p)$ and will either be real negative numbers or occur in complex conjugate pairs of negative real part.

For many purposes it is more convenient to introduce the logarithmic gain, defined as

$$\chi(p) = \log G(p) = \sum_{r=1}^{N_1} \log(p - p_r) - \sum_{s=1}^{N_2} \log(p - q_s) \quad (2)$$

In terms of χ , the impulse response $A(t)$ becomes

$$A(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \epsilon^{pt+\chi(p)} dp \quad (3)$$

If circuits are cascaded in the manner mentioned earlier, the impulse response of n stages is

$$A_n(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \epsilon^{pt} G^n(p) dp = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \epsilon^{pt+n\chi(p)} dp \quad (4)$$

If G is a rational function of the type considered, it is permissible to distort the contour so long as all singularities lie to its left. It is also permissible to complete it with a large semi-circle in the left half-plane. If the contour is made a circle on the origin outside all singularities and G is expanded in power series of $1/p$, the impulse response for small values of t is given by the equation

$$A_n(t) \simeq G_0^n \frac{t^{nN-1}}{(nN-1)!}$$

Hence the response per unit gain will not become appreciable until times of the order of magnitude of n are reached.

(2.2) Evaluation in Terms of Real Frequencies

The problem is therefore to evaluate the integral, eqn. (4), for large values of t and n , t being measured in some unit convenient for the network. Taking the contour along the real frequency axis and writing $\chi(j\omega) = \alpha(\omega) + j\beta(\omega)$, eqn. (4) becomes

$$\begin{aligned} A_n(t) &= \frac{1}{2\pi} \int_0^\infty \exp \{ j\omega t + n[\alpha(\omega) + j\beta(\omega)] \} d\omega \\ &= \frac{1}{\pi} \int_0^\infty \epsilon^{n\alpha} \cos(\omega t + n\beta) d\omega \end{aligned}$$

Numerical integration is possible, but tedious because of the rapid oscillations of the integrand. To obtain an approximate analytical expression, a method devised by Kelvin may be used. A modern enunciation of this method is given by Watson.⁴ This is the method employed by Carson,² referred to above. A more powerful tool, known as the method of steepest descent, can be used to transform the integral, eqn. (4), into a form more suited for computation and approximation.

(2.3) Application of the Method of Steepest Descent

The integral of eqn. (4) may be written

$$A_n(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \epsilon^{nF(p)} dp$$

where $F = \frac{pt}{n} + \chi$.

The method of steepest descent, described in various text books,^{5,6} consists in changing the contour of integration into one upon which $\mathcal{I}(F)$ is constant, and $\mathcal{R}(F)$ has a maximum at one point of the contour, decreasing as rapidly as possible away from this point. In particular cases it may be necessary to reform the contour into several portions, each having a maximum point. Because of the nature of the surface obtained by plotting $\mathcal{R}[F(p)]$ against real and imaginary parts of p , these maximum points are called *cols*, and their positions are determined by the equation $dF/dp = 0$.

Assuming that a suitable contour can be found, the contribution of the part near one col is

$$\int \epsilon^{nF} dp = \epsilon^{n(\phi_c + j\psi_c)} \int \epsilon^{n(\phi - \phi_c)} dp$$

where $F = \phi + j\psi$, and the suffix c denotes values at the col. The quantity $\phi - \phi_c$ is always negative, and hence the integrand may be written $\epsilon^{-n\sigma}$. For a given value of σ , the parts of the contour away from the col contribute less to the value of the integral as n becomes larger. Hence an estimate of the integral can be obtained in terms of the behaviour near the col.

The work which follows is closely parallel to that of Watson on Bessel functions.⁷

(2.4) Parallel RC Stages

In order to decide whether the method can be applied to the type of function involved in eqn. (4), it is profitable to consider simple cases; the simplest possible is that of a valve with resistive anode load shunted by capacitance.

Here $G(p) = g/C \left(p + \frac{1}{RC} \right)^{-1}$

$$= G_0(pRC + 1)^{-1}$$

$$\begin{aligned} A_n(t) &= \frac{G_0^n}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\epsilon^{pt}}{(pRC + 1)^n} dp \\ &= \frac{G_0^n}{2\pi j} \frac{1}{RC} \int_{c-j\infty}^{c+j\infty} \frac{\epsilon^{z(t/RC)}}{(z + 1)^n} dz \quad (5) \end{aligned}$$

This can be evaluated exactly, and gives

$$A_n(t) = \frac{G_0^n}{RC} \frac{1}{(n-1)!} \left(\frac{t}{RC} \right)^{n-1} \epsilon^{-t/RC}$$

If $\tau = t/RC$,

$$A_n(\tau) = \frac{G_0^n}{RC} \frac{\tau^{n-1}}{(n-1)!} \epsilon^{-\tau} \quad (6)$$

To apply the method of steepest descent, eqn. (5) is written in the form

$$RCG_0^{-n} A_n(\tau) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \epsilon^{nF} dz \quad (7)$$

where

$$F = \frac{p\tau}{n} + \chi = \frac{z\tau}{n} - \log(z + 1) \quad (8)$$

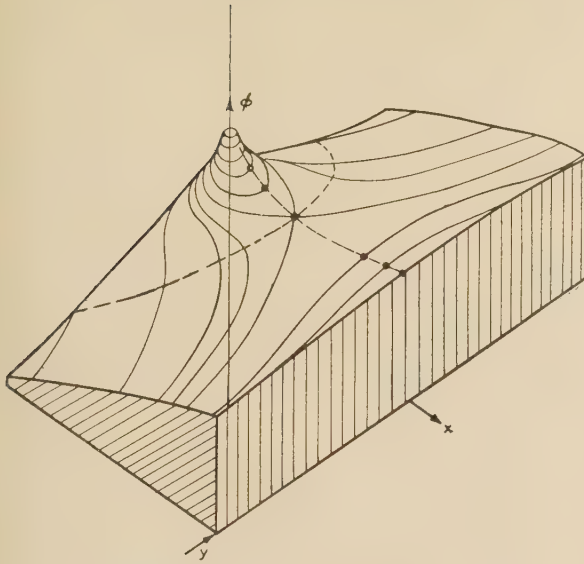


Fig. 1.—Relief map of $\mathcal{R}\left[\frac{z\tau}{n} - \log(z+1)\right]$.

— $\phi = \mathcal{R}(F) = \text{constant}$.
 - - - - Lines of steepest ascent or descent.

A relief plot of the surface $\mathcal{R}(F)$ is shown in Fig. 1. The position of the col is given by the equation $dF/dz = 0$, which on substitution from eqn. (8) gives

$$z_c = -1 + \frac{1}{\lambda} \quad (9)$$

where λ has been written for τ/n . Also

$$\left. \begin{aligned} F_c &= 1 - \lambda + \log \lambda \\ \sigma &= 1 - \lambda(z+1) + \log \lambda(z+1) \end{aligned} \right\} \quad (10)$$

The contour through the point $z = z_c$ is given by the equation $\mathcal{R}(F) = 0$, and is of the form shown in Fig. 1, tending asymptotically to the lines $y = \pm \pi/\lambda$ as $x \rightarrow -\infty$. The considerations of Section 2.1 show that the contour $c \pm j\infty$ may be replaced by this contour, C .

Hence we may write

$$RCG_0^{-n} A_n(\tau) = \frac{\varepsilon^{nF_c}}{2\pi j} \int_C \varepsilon^{-n\sigma} dz \quad (11)$$

It may be observed that there is nothing approximate about this expression, and that given a method of finding the contour and σ , it is ideal for numerical computation. This matter will be again referred to in Section 3. To obtain an approximate expression for the result of the integration, the procedure is to change the variable of integration to σ , when the range of integration is zero to infinity. To do this $dz/d\sigma$ must be found. σ is known as a function of z from eqns. (10), and hence as a function of $\zeta = z - z_c$. Since σ vanishes for $\zeta = 0$, expansion for σ in powers of ζ may be obtained. This series may be inverted to give ζ in terms of σ , which will by symmetry be a two-valued function. If the two values of z are z_2, z_1 , eqn. (11) can be written

$$RCG_0^{-n} A_n(\tau) = \frac{\varepsilon^{nF_c}}{2\pi j} \int_0^\infty \varepsilon^{-n\sigma} \left(\frac{dz_2}{d\sigma} - \frac{dz_1}{d\sigma} \right) d\sigma \quad (12)$$

with the integrand known, as a function of σ . It is shown, in

Reference 7, that if a term-by-term integration is carried out, an asymptotic series⁵ is obtained.

The algebra of the process outlined above is performed in Section 12.1, and yields the result

$$RCG_0^{-n} A_n(\tau) \sim \frac{1}{\sqrt{2\pi}} \varepsilon^{n-\tau} \tau^{n-1} n^{-n+1/2} \left(1 - \frac{1}{12n} + \dots \right) \quad (13)$$

The application of Stirling's approximation to $(n-1)!$ in the exact expression of eqn. (6) yields an identical expression. It is evident, therefore, that the method gives a correct answer to the problem.

For any form of network which has a pole on the real-frequency axis, the general ideas follow very closely on those for the single pole. To consider the effects of under-damped systems, another ideal problem must be chosen. Two poles at conjugate complex points form a suitable configuration to bring out the salient features.

(2.5) The Case of Two Poles

In terms of a reduced variable z of the type considered in Section 2.4, the network function may be chosen to be

$$\chi = -\log(z^2 + 1) \quad (14)$$

This has singularities at the points $\pm j$. Although this is physically unrealizable as it stands, the effect of the damping displacing the singularities to $\pm j - \alpha$ is merely to multiply the impulse response by $\varepsilon^{-\alpha t}$, so that no generality is lost by taking the expression of eqn. (14).

The impulse response is

$$A_n(\tau) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{(z^2 + 1)^n} \varepsilon^{z\tau} dz \quad (15)$$

This can be evaluated exactly, giving

$$A_n(\tau) = \frac{\sqrt{\pi}}{(n-1)!} \left(\frac{\tau}{2} \right)^{n-1/2} J_{n-1/2}(\tau) \quad (16)$$

Since this case requires different treatment from that in Section 2.4, only a general discussion of the type of integral obtained by the application of the method of steepest descent is given in this Section.

In Fig. 2 are shown, for different values of τ/n , contours of $\mathcal{R}(F)$ constant and lines of steepest descent and ascent for $F = \tau z/n - \log(z^2 + 1)$. Solution of the equation $F' = 0$ shows that for $\tau/n < 1$, cols exist at the points $\frac{1}{\lambda} \pm \sqrt{\left(\frac{1}{\lambda^2} - 1 \right)}$,

of which the one with the positive sign is relevant to the contour integration. For $\tau/n > 1$ the cols appear at points $\frac{1}{\lambda} \pm j\sqrt{\left(1 - \frac{1}{\lambda^2} \right)}$, which lie on the unit circle. This is in contrast to the case of Section 2.4, where a col always existed on the X -axis. For a col to exist on the real axis, $\tau/n = -X'$ must be satisfied. In Section 2.4, $-X'$ increases from 0 to $+\infty$ as x changes from $+\infty$ to -1 . In the present case X' has a minimum value of -1 occurring at $x = 1$, and hence cols on the real axis are possible only for $\tau/n < 1$.

The contours in this case are either like that of Section 2.4 or consist of two sections, as shown in Fig. 2(c), which are mirror images in the X -axis.

From these results, the contours likely to be obtained for a more general network can be considered.

(2.6) A General Network

Although it is difficult to draw a completely general picture for any distribution of poles and zeros, the distribution of

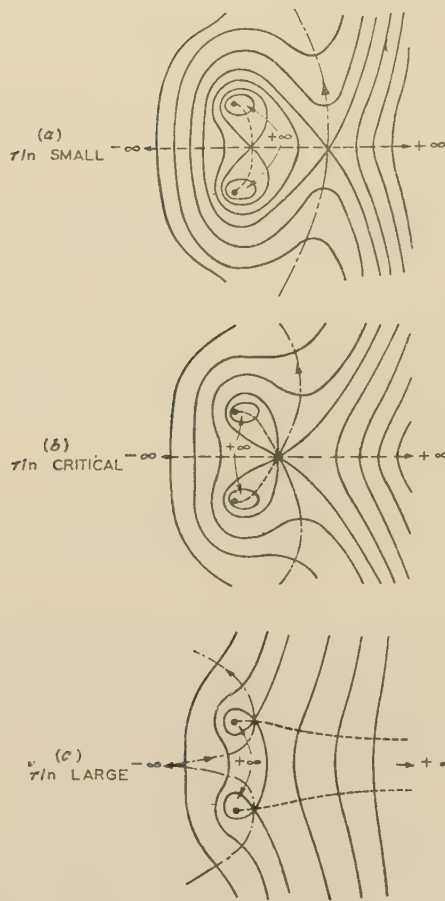


Fig. 2.—Approximate contours for $F = \frac{z\tau}{n} - \log(z^2 + 1)$.

— $\Re(F) = \text{constant}$.
 - - - Contour.
 . . . Other lines of steepest ascent or descent.

relevant cols and the appropriate contours is clear for τ/n either small or large. In the former case there will be a col on the real axis a long way from the centre of the pole-zero distribution, and a contour of the type discussed in Section 2.4 will exist. When τ/n is large there will be a col in the immediate neighbourhood of each pole, with contours of the type illustrated in Fig. 2(c) for each pole.

The mode of change from one pattern to the other must be studied for each distribution. One common case is very like that of the two poles in Section 2.5, i.e. when a minimum of $d\chi/dz$ exists along the positive real axis. This will always be true if no pole exists on the real axis.

It is evident that two cases need be considered from which to build up the response once the contour behaviour is known. These are (a) a path passing over a single col on the real axis, and (b) two segments of a path each relating to a col of a conjugate complex pair. On each path or segment the real part of $\lambda z + \chi$ will increase from negative infinity through a maximum back to negative infinity. Expressions for the contour integral of eqn. (4) appropriate to each of these cases will now be obtained.

(2.6.1) Single Col on the Real Axis.

The path to be considered is shown as C' in Fig. 3. As before, let $F = \lambda z + \chi$, and $\sigma = F_c - F$. Along C' , σ will be real, and for each value there will be two points z_1 and z_2 on the

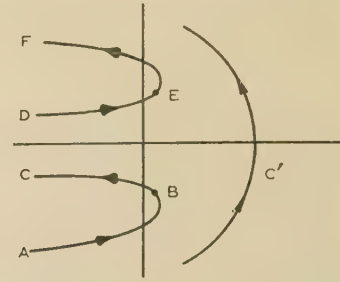


Fig. 3.—Contours of the inversion integral.

contour, one in the lower half-plane and the other in the upper. Hence the contour integral becomes

$$A_n(\tau) = \frac{\varepsilon^{nF_c}}{2\pi j} \int_0^\infty \varepsilon^{-n\sigma} \left(\frac{dz_2}{d\sigma} - \frac{dz_1}{d\sigma} \right) d\sigma \quad (18)$$

(2.6.2) Cols of a Conjugate Complex Pair.

The segments of the path are now ABC, DEF in Fig. 3. If σ denote $F'_c - F$ and $F''_c - F$ for the lower and upper half-planes respectively, then for each value of σ on the contour four values of z will exist. Let these be z_1, z_2, z_3, z_4 in AB, BC, DE, EF respectively. By symmetry $F'_c = F''_c^*$, $z_1 = z_4^*$, $z_2 = z_3^*$. Hence the contour integral becomes

$$\begin{aligned} A_n(\tau) &= \frac{\varepsilon^{nF'_c}}{2\pi j} \int_{ABC} \varepsilon^{-n\sigma} dz + \frac{\varepsilon^{nF''_c}}{2\pi j} \int_{DEF} \varepsilon^{-n\sigma} dz \\ &= \frac{\varepsilon^{nF'_c}}{2\pi j} \int_0^\infty \varepsilon^{-n\sigma} \left(\frac{dz_2}{d\sigma} - \frac{dz_1}{d\sigma} \right) d\sigma + \frac{\varepsilon^{nF''_c}}{2\pi j} \int_0^\infty \varepsilon^{-n\sigma} \left(\frac{dz_4}{d\sigma} - \frac{dz_3}{d\sigma} \right) d\sigma \\ &= \frac{1}{\pi} \left[\varepsilon^{nF'_c} \int_0^\infty \varepsilon^{-n\sigma} \left(\frac{dz_4}{d\sigma} - \frac{dz_3}{d\sigma} \right) d\sigma \right] \quad (19) \end{aligned}$$

(2.6.3) Response Associated with Type of Contour.

Consider the form of eqn. (18). F_c is real, and as λ changes will increase steadily as long as the col exists on the real axis. The value of the integral would also be expected to change slowly. Hence the impulse response would be expected to increase slowly as long as it is described by an integral of this form. Once the form of eqn. (19) applies, however, rapid variations can be expected, since it is required to find the imaginary part of an expression containing $\exp(nF'_c)$. Hence the value of λ when the form of contour changes assumes importance as a criterion for the time when oscillations of the impulse response can be expected. For the case when a minimum of $d\chi/dz$ occurs in the real axis, this critical value of λ is the greatest value for which the equation

$$\lambda + \frac{d\chi}{dz} = 0 \quad (20)$$

has a real root. This will occur at the largest positive real root of $d^2\chi/dz^2 = 0$. In cases when a pole exists on the real axis, there will still be a value of λ for which the contour changes from one passing over a single col to two or more separate contours passing over cols at conjugate complex positions or on the real axis. In such cases the critical value of λ is no longer determined by a simple condition.

In order to obtain approximate expressions, three regions of values of λ must be considered separately: $\lambda \ll 1$, $\lambda \gg 1$ and λ

near the critical value. It is to be expected that the point of change-over from one form to the other will occur at a distance from the origin of the order of the distance of the singularities from the origin. Hence the critical value of λ will be of the order of unity.

(3) POSSIBILITY OF EXACT EVALUATION OF THE IMPULSE RESPONSE BY ANALOGUE COMPUTING

The impulse response as given by eqns. (18) and (19) is exact, and, as mentioned earlier, is in the most suitable form for numerical integration. The difficulty in evaluation is that both the contour of integration and the functions change with each value of the time taken. It is quite possible to obtain the data necessary for computation by means of the electrolytic-tank realization of the electrostatic analogy.⁸ This analogue has been used by Boothroyd *et al.*⁹ to obtain transient responses, but their method is easily applicable only to the case of simple poles, and not when cascaded networks are considered.

The data to be determined before computation are

- The position of the relevant col or cols for each value of time.
- The values of the real and imaginary parts of F at the col.
- The variation of the real part of F along the contour.

Once this information has been obtained for a given network, the numerical evaluation of eqns. (18) and (19) has to be done, but this is easy because of the nature of the integrand. If the impulse response for one case only is required the method is laborious, but if several values of n are required it is possible that labour will be saved. It may be observed that the integrand considered in Section 2.2 decreases as an inverse power of ω , whereas in eqns. (18) and (19) the decrease is exponential.

(4) ASYMPTOTIC EXPRESSIONS FOR THE IMPULSE RESPONSE

(4.1) Times for which λ is much less than the Critical Value

For values of λ less than critical, the position of the col is on the real axis, and the contour is like that in Section 2.4. The network function X may be expanded in the form

$$X = -N \left(\log z + \frac{A_1}{z} + \frac{A_2}{z^2} + \dots \right)$$

With the aid of this expansion the equation $\lambda + X' = 0$ can be formulated, and a solution for the position of the col obtained in the form of a series in powers of λ/N . From this point the method of solution follows exactly as in Section 2.4. The first term of the series obtained gives

$$A_n(\tau) \sim \frac{1}{(nN-1)!} \tau^{nN-1} e^{-A_1 \tau} \quad (21)$$

in which the factorial has been written in place of an expression essentially equal to it for large values of nN . The series is in ascending powers of λ/N and is therefore useless unless λ/N is small; i.e. since N is a small integer, it fails for λ near the critical value. It therefore describes the slow build-up, and, of course, the result is familiar.

(4.2) Times for which λ is much greater than the Critical Value

For values of λ greater than critical the configuration is as described in Section 2.6.2, leading to eqn. (19). The process of finding the functions in the integrand is done by taking an

expansion for G valid in the neighbourhood of a pole at z_1 , near which a col will lie. This expansion is

$$G = \frac{G_1}{z - z_1} [1 + a_1(z - z_1) + a_2(z - z_1)^2 + \dots]$$

The first term of the result obtained is

$$A_n(\tau) \sim \frac{2|G_1|^n}{(n-1)!} \tau^{n-1} e^{\tau x_1} \cos(\tau y + n \arg G_1) \quad (22)$$

This again is what is to be expected. The contribution to the impulse response of each pair of poles is a damped oscillation of frequency and decay given by the position of the pole.

Neither of these two results is new, but they serve to show that the method can give valid results. Interest centres chiefly in the rate of build-up when the response is becoming appreciable and in the subsequent behaviour. All this occurs in the range of λ near the critical value.

(4.3) Times for which λ is near the Critical Value

The present paper deals only with the case when the critical value is determined by the vanishing of the second derivative of X . It will appear that information on the response in this neighbourhood is difficult to obtain by the methods of Sections 4.1 and 4.2, and better results are obtained by the method given in Section 5. The value of the response at the critical value of λ can be simply obtained in the form of an asymptotic expansion. The process differs slightly from the previous ones, since at the critical point the second derivative of X vanishes.

The result obtained is

$$\begin{aligned} A_n(\tau_0) &\sim \frac{e^{nF_0}}{2\pi\sqrt{3}} \left[\left(\frac{6}{nX_0'''} \right)^{1/3} \Gamma(1/3) - \frac{1}{6} \frac{X_0^{iv}}{X_0'''} \left(\frac{6}{nX_0'''} \right)^{2/3} \Gamma(2/3) + \dots \right] \\ &\sim e^{nF_0} \left[\frac{0.4473}{(nX_0''')^{1/3}} - \frac{X_0^{iv}}{X_0'''} \frac{0.0685}{(nX_0''')^{2/3}} + \dots \right] \quad (23) \end{aligned}$$

where the suffix 0 denotes values at the critical point.

It may be observed that, since the critical point occurs for the greatest zero of X'' other than infinity, there must be a maximum of X''' between the critical point and infinity on the real p axis. At this maximum X'''' will vanish, and hence at the critical point X_0''' will be opposite in sign to X'''' as $p \rightarrow +\infty$. Hence X_0''' will be positive. For values of n about 10, these two terms give results accurate to within about 1% in normal cases.

Further investigation by these methods is difficult. A series valid for λ different from the critical value by less than $n^{1/3}$ can be obtained, but this is not of any practical use. A less rigorous method, again essentially as described by Watson,⁷ which gives results identical to the previous results in the appropriate ranges of λ , can also give information of value about the behaviour in the interesting region. This is described in Section 5. It has the disadvantage that whereas operations given in the previous Sections produce results by a definite known process, so that errors can be estimated, the method to be described can only be tested by taking examples for which exact solutions are known.

(5) AN APPROXIMATE METHOD

This method depends on the fact, mentioned in Section 2.3, that the chief contribution to the integrals of eqns. (18) and (19) comes from a region near the col over which the contour passes. The larger the value of n , the more restricted is this region. Hence an approximation to the required integral should be obtained by taking as integrand a function approximating to

the actual integrand near the col, and integrating it over a contour the same as the actual contour near the col. Any result obtained by this process must be checked against the known results expressed by eqns. (21), (22) and (23).

An obvious choice for the approximating function is the first terms of the power series expansion of σ about the col, i.e.

$$\sigma \simeq \bar{\sigma} = -\frac{\chi_c''}{2!}(z - z_c)^2 - \frac{\chi_c'''}{3!}(z - z_c)^3 \quad (24)$$

Eqns. (18) and (19) may be written in the form

$$A_n(\tau) = \frac{\varepsilon^{nF_c}}{2\pi j} \int_C \varepsilon^{-n\sigma} dz \quad (25)$$

$$A_n(\tau) = \frac{1}{\pi} \oint \left(\varepsilon^{nF_c} \int_C \varepsilon^{-n\sigma} dz \right) \quad (26)$$

The contour in both cases is one for which σ is real and positive. Hence an approximation to the integral parts should be obtained

by evaluating $\int_{C'} \varepsilon^{-n\sigma} dz$, C' being a contour on which $\bar{\sigma}$ is real and positive. The investigation falls into two parts, λ greater and less than λ_0 , the critical value.

(5.1) Times for which λ is less than the Critical Value

The function $\bar{\sigma}$ can be written

$$\bar{\sigma} = -\alpha_1 \zeta^2 - \alpha_2 \zeta^3 \quad (27)$$

where $\zeta = z - z_c$, $\alpha_1 = \chi_c''/2$, $\alpha_2 = \chi_c'''/6$. The values of α_1 and α_2 are real and positive in some region with $\lambda \leq \lambda_0$, although α_2 eventually becomes negative. By previous considerations the single col lies on the real axis.

Using the result obtained in Section 12.2, the approximate expression for $A_n(\tau)$ becomes, from eqn. (25),

$$A_n(\tau) \simeq \frac{1}{(3n|\alpha_2|)^{1/3}} [\exp(nF_c + \frac{2}{3}\mu^{3/2}) \text{Ai}(\mu)] \quad (28)$$

in which $\mu = \frac{n^{2/3}\alpha_1^2}{(3|\alpha_2|)^{4/3}}$, and $\text{Ai}(\mu)$ is an Airy function. For small values of τ it may be shown that eqn. (28) reduces to eqn. (21), and for the critical value to the first term of eqn. (23).

(5.2) Times for which λ is greater than the Critical Value

The function $\bar{\sigma}$ is as in eqn. (27), but α_1 and α_2 are now complex. In Section 12.3 it is shown that the contribution to $A_n(\tau)$ due to one pair of poles is given by

$$A_n(\tau) = \mathcal{R} \left[\frac{\exp(nF_c + \frac{2}{3}\mu^{3/2})}{k(3n\alpha_2)^{1/3}} h_1(-\mu) \right] \quad (29)$$

where $k = 12^{1/6} e^{-j\pi/6}$, $\mu = \frac{n^{2/3}\alpha_1^2}{(3\alpha_2)^{4/3}}$

and $h_1(-\mu)$ is a modified Hankel function.

It may be shown that this function gives the first terms of the approximations for the critical point and for τ large, as were obtained previously.

(5.3) Practical Use of the Results of Sections 5.1 and 5.2

The approximations will evidently be only of use when n is large enough to make the region near each col the only one of importance in the evaluation of the integrals. It is difficult to obtain a close estimate of the errors involved by analytical means,

and it seems easiest to resort to particular cases which can be evaluated exactly. Two of these have been used as check problems. They are (a) the Wheeler¹⁰ coupling network, which involves the characteristic impedance of a low-pass filter, and hence an infinite number of components, and (b) the case of two poles discussed in Section 2.5. These will be considered in detail in the following Sections.

(6) EVALUATION OF PROBLEMS WITH KNOWN SOLUTIONS

(6.1) The Wheeler Coupling Network

In essence the Wheeler circuit is shown in Fig. 4. If the

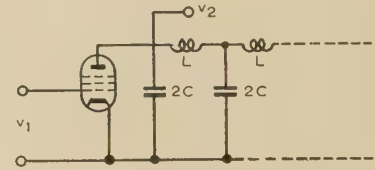


Fig. 4.—Circuit for Wheeler coupling network.

characteristic impedance at zero frequency is R_0 , the impedance Z in the anode circuit of the pentode is given by

$$\frac{1}{Z} = pC + \frac{1}{R_0} \sqrt{\left(\frac{p^2}{\omega_0^2} + 1\right)}$$

and, since $\omega_0 CR_0 = 1$,

$$\frac{1}{Z} = \frac{1}{R_0} \left[\frac{p}{\omega_0} + \sqrt{\left(\frac{p^2}{\omega_0^2} + 1\right)} \right]$$

Hence, when normalized, $G = [z + \sqrt{z^2 + 1}]^{-1}$

For n stages of this network, the impulse response is given by

$$A_n(\tau) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\varepsilon^{z\tau}}{[z + \sqrt{z^2 + 1}]^n} dz$$

Reference to a table of Laplace transforms shows that this gives

$$A_n(\tau) = \frac{n}{\tau} J_n(\tau) \quad (30)$$

In Fig. 5 the lines are calculated from eqn. (30), for $n = 6, 10$

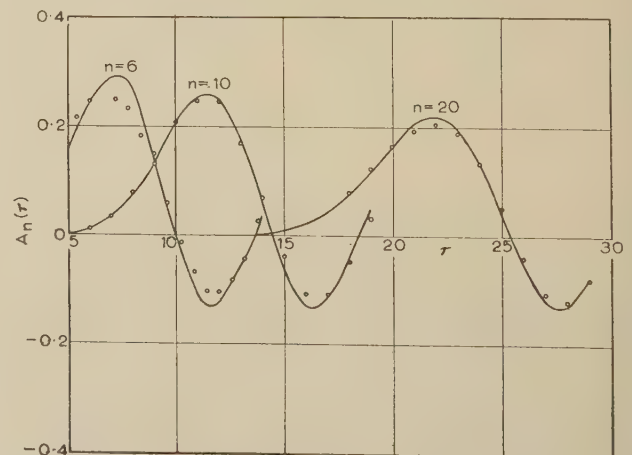


Fig. 5.—Impulse responses for circuit shown in Fig. 4.

— Exact.
○ Approximate.

and 20, and the points are those obtained from eqns. (28) and (29).

The greatest error occurs for $n = 6$, as expected. The errors for $n = 10$ appear to be less than for $n = 20$, which is slightly unexpected. However, even though the errors are expected to tend to zero as n becomes very large, there is no guarantee that the approach is monotonic. Apart from those for $n = 6$, the errors are of the order of 10% of maximum excursion.

(6.2) Coupling Circuit Characterized by Two Poles

This type of circuit has been briefly discussed in Section 2.5. A circuit which can have this type of network function for suitable component values is shown in Fig. 6. To obtain a

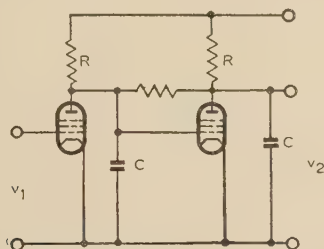


Fig. 6.—Coupling circuit characterized by two poles.

realistic impulse response the poles have been taken at the points $-1 \pm j$. The network function G then becomes $[(z + 1)^2 + 1]^{-1}$, normalized to unity gain at low frequencies. Modifications to eqns. (15) and (16) then give the exact impulse response as

$$A_n(\tau) = \frac{2^n \sqrt{\pi}}{(n-1)!} \left(\frac{\tau}{2}\right)^{n-1/2} \varepsilon^{-\tau} J_{n-1/2}(\tau)$$

This is plotted in full lines in Fig. 7, for three values of n over ranges dictated by the tables available.

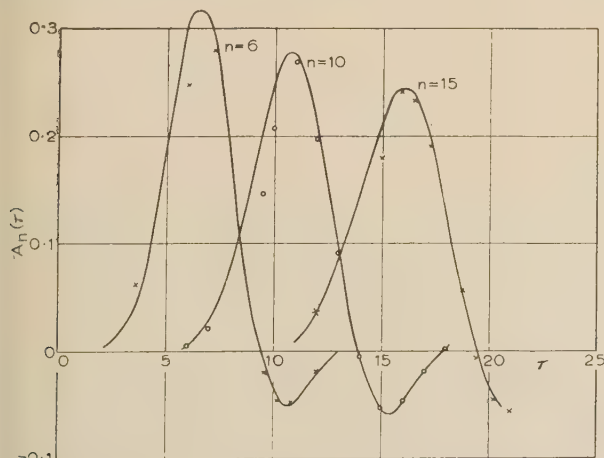


Fig. 7.—Impulse responses for circuit shown in Fig. 6.

— Exact.
× ○ Approximate.

The results of calculations from eqns. (28) and (29) are shown as the points in Fig. 7. It will be observed that relatively large errors occur on the initial slope, but that after the maximum the approximation is very good, even for $n = 6$. The error near the first peak is mathematically due to the neglect of the fourth derivative of X . This neglect might be expected to lead to the most serious errors near the critical point, since the second derivative vanishes at this point. This error is absent in the

case of Section 6.1, where the fourth derivative at the critical point is zero. It will be found that insertion of the second term of eqn. (23) gives a much higher degree of accuracy. It is possible that a correcting term could be found for use in such cases. Notwithstanding the absolute error, the height of the first maximum is given to within about 2%, although slightly displaced in time. The response after the maximum is given virtually exactly by the approximate expression for $n > 10$.

(7) A FURTHER APPROXIMATION VALID FOR LARGE VALUES OF n

In this Section a simplified version of the results of Section 5 is presented, which might be expected to yield results in the neighbourhood of the critical point.

In eqn. (29) the quantities F_c , μ , and α_2 can be regarded as functions of λ . It can be shown that as λ is varied near the critical point, both $\mu^{1/2}$ and the position of the col may be determined in the form of a power series in $(\lambda - \lambda_0)^{1/2}$. Hence F_c can also be expressed in a power series. If derivatives higher than the third at the critical point are neglected, it is found that

$$z_c - z_0 = j \left(\frac{2}{X_0'''} \right)^{1/2} (\lambda - \lambda_0)^{1/2} + O(\lambda - \lambda_0) \quad (31)$$

$$\mu = -n^{2/3} \left(\frac{2}{X_0'''} \right)^{1/3} (\lambda - \lambda_0) + O[(\lambda - \lambda_0)^{3/2}] \quad (32)$$

$$nF_c + \frac{2}{3}\mu^{3/2} = nF_0 + n(\lambda - \lambda_0)z_0 + O[(\lambda - \lambda_0)^2] \quad (33)$$

An approximate expression for that part of $A_n(\tau)$ due to one pair of cols only, valid for $\lambda - \lambda_0$ small, is obtained by substituting these values in eqn. (29). Retaining only the lowest terms it is found that

$$A_n(\tau) \simeq \varepsilon^{nF_0 + n(\lambda - \lambda_0)z_0} \left(\frac{2}{nX_0'''} \right)^{1/3} \mathcal{R} \left\{ \frac{1}{k} h_1 \left[n^{2/3} \left(\frac{2}{X_0'''} \right)^{1/3} (\lambda - \lambda_0) \right] \right\} \quad (34)$$

Since the argument of the Hankel function is real, this is more conveniently expressed by means of the Airy function in the form

$$A_n(\tau) \simeq \varepsilon^{nF_0 + n(\lambda - \lambda_0)z_0} \left(\frac{2}{nX_0'''} \right)^{1/3} \text{Ai} \left[-n^{2/3} \left(\frac{2}{X_0'''} \right)^{1/3} (\lambda - \lambda_0) \right] \quad (35)$$

It can also be shown that this same equation applies to the case $\lambda < \lambda_0$. A graph of the Airy function for real argument is shown in Fig. 8.

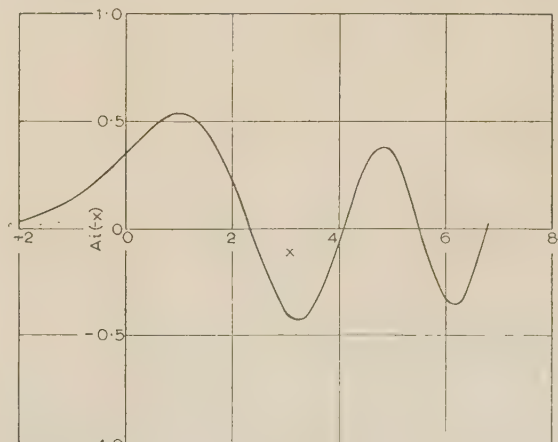


Fig. 8.—Curve of $\text{Ai}(-x)$.

It is evident that, for example, the first zero of this term is determined by the argument of the Airy function reaching a certain value. As n increases, $\lambda - \lambda_0$ will decrease to maintain this fixed value, i.e. $\lambda - \lambda_0$ will decrease as $n^{-2/3}$ for points in the neighbourhood of the build-up. Hence the expansion, eqn. (35), should give an approximation for sufficiently large values of n . Since, however, the terms decrease as $n^{-1/3}$, it needs a very large value of n to obtain high accuracy. It is nevertheless worth while investigating some of the results to be obtained from eqn. (35), since its form is much simpler than the more accurate expression (29). In assessing results obtained, it is important to realize that only the value of a function and its first three derivatives at a single point are required. This is meagre information compared, for example, with that contained in a gain/frequency curve.

(7.1) The First Zero of the Impulse Response

If, near the critical point, the contour of the integral passes over one pair of cols only, as will be usual for a certain range of λ , eqn. (35) is a complete expression for the impulse response. The first zero will correspond to the first zero of the Airy function, which occurs when its argument is -2.34 . Hence the time of the first zero is given approximately by

$$\lambda - \lambda_0 = 2.34n^{-2/3} \left(\frac{\chi_0'''}{2} \right)^{1/3}$$

or
$$\tau = n\lambda_0 + 1.86(n\chi_0''')^{1/3} \quad (36)$$

The form of this expression can be compared with that obtained in the two particular cases described in Section 6. In Section 6.1 the zeros of the impulse response are those of $J_n(\tau)$, and in Section 6.2 those of $J_{n-1/2}(\tau)$. It is known¹¹ that the first zero of the Bessel function $J_\nu(x)$ is given approximately by

$$\nu + 1.8557\nu^{1/3} + O(\nu^{-1/3}) \quad (37)$$

For both cases $\lambda_0 = \chi_0''' = 1$, whence

$$\tau = n + 1.86n^{1/3} \quad (38)$$

This is accurate within $O(n^{-1/3})$ for Section 6.1, putting $\nu = n$ in eqn. (37), but for Section 6.2 eqn. (37) gives

$$\tau = n - \frac{1}{2} + 1.8557(n - \frac{1}{2})^{1/3}$$

i.e. the error in eqn. (38) is $O(1)$ and not $O(n^{-1/3})$. However, for $n = 10$ the error is less than 4%.

A further test of the accuracy of eqn. (36) has been made by comparison with results obtained for inductance-compensated pentodes (Fig. 9) by Bedford and Fredendall.¹² In that paper

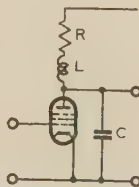


Fig. 9.—The inductance-compensated amplifier.

the response to a step function was calculated for a number of different compensation conditions for numbers of stages up to 64. The first maximum of this response corresponds to the first zero of the impulse response. The parameters of the circuit required to make the approximate calculations are evaluated in Section 12.4. Two sets of comparison have been made, one for circuit parameters giving a large overshoot by the time 64 stages

have been reached, and the other for parameters giving only a small overshoot. The results obtained are shown in Table 1.

Table 1

TIME OF FIRST ZERO OF IMPULSE RESPONSE (UNITS OF $2\pi RC$)

K	Number of stages					Remarks
	4	8	16	32	64	
1.41	0.72	1.16	1.96	3.47	6.38	Calculated from eqn. (36)
	0.74	1.14	1.95	3.45	6.35	Bedford and Fredendall
1.61			2.23	3.97	7.31	Calculated from eqn. (36)
			2.25	4.00	7.33	Bedford and Fredendall

The agreement between the two sets of results is remarkable. The differences are, in fact, less than the errors in reading the results from the curves in Bedford and Fredendall's paper. It therefore appears that eqn. (36) is a good guide to the position of the first maximum in the response to a step function of this type of video-amplifier compensation circuit.

(7.2) Estimation of Circuit Values for a Given Small Overshoot in the Response to a Step Function

Eqn. (35) can be used to get an idea of the type of response to be expected for a step input. If attention is restricted to the type of circuits considered in Section 7.1, this equation is a complete approximation, since only two cols will be involved.

If in eqn. (35) the argument of the Airy function is denoted by $-x$, we have

$$A_n(\tau) \simeq \left(\frac{2}{n\chi_0'''} \right)^{1/3} \varepsilon^{nF_0} \exp \left[xz_0 \left(\frac{n\chi_0'''}{2} \right)^{1/3} \right] \text{Ai}(-x) \quad (39)$$

where

$$x = n^{2/3} \left(\frac{2}{\chi_0'''} \right)^{1/3} (\lambda - \lambda_0)$$

As n varies, there will be an 'amplitude factor' $\left(\frac{2}{n\chi_0'''} \right)^{1/3} \varepsilon^{nF_0}$ which will primarily depend on F_0 for its behaviour, and a 'form factor' $\exp \left[xz_0 \left(\frac{n\chi_0'''}{2} \right)^{1/3} \right]$. The coefficient χ_0''' is always positive; the signs of F_0 and z_0 depend on the circuit. If z_0 is positive, the maxima of eqn. (39) will tend to move up towards the zeros of $\text{Ai}(-x)$, and the successive maxima and minima of eqn. (39) will increase. In terms of response to a step, which will be the integral of eqn. (39), this means that the rings will tend to increase in amplitude, and that the falling part between maximum and minimum will be steeper than the rising part. If F_0 is positive a further increase will take place on account of the amplitude factor. If F_0 is negative, as n is increased indefinitely the rings will be reduced in amplitude but the form will be unchanged, since n increases more rapidly than $n^{1/3}$. If z_0 is negative, successive maxima and minima of the impulse response will decrease in amplitude, and the maxima will tend to move backwards away from the following zero. In terms of the response to a step, the rings will die away. The amplitude factor again comes in to determine scale.

An estimate of the ring may be simply obtained. If the gain

has been normalized to unity at low frequencies, the final value of the response to a step will be unity. The difference between the first maximum and minimum of the step response will be given by the integral of eqn. (39) with respect to τ between the first and second zeros of $A_i(-x)$. Since an approximate answer only is required—and usually it will be required that the difference be small—refined integration is not called for. An estimate may be made by regarding $A_n(\tau)$ as parabolic and taking an approximate peak height given by its value at the minimum of $A_i(-x)$. Reference to Fig. 8 shows that the first and second zeros of $A_i(-x)$ are 2.34 and 4.09, and that its first minimum is at -3.25 , magnitude -0.42 . Hence the difference between first and second ring in the step response is

$$\begin{aligned} \int_{\tau_1}^{\tau_2} A_n(\tau) d\tau &= \varepsilon^{nF_0} \int_{x_1}^{x_2} \exp \left[x z_0 \left(\frac{nX_0'''}{2} \right)^{1/3} \right] A_i(-x) dx \\ &\simeq -\varepsilon^{nF_0} \frac{2}{3} \exp \left[3 \cdot 3 z_0 \left(\frac{nX_0'''}{2} \right)^{1/3} \right] \times 0.42 \times 1.75 \\ &\simeq -\frac{1}{2} \exp \left[nF_0 + 3 \cdot 3 z_0 \left(\frac{nX_0'''}{2} \right)^{1/3} \right] \end{aligned} \quad (40)$$

This expression represents the ratio of difference to final amplitude. If this ratio is to be restricted, as in a video amplifier, to, say, 5%, an equation is obtained which can be used to determine a suitable value for a parameter determining F_0 and z_0 . The results obtained by this procedure have been investigated for two cases, the inductance-compensated amplifier, and the circuit considered in Section 6.2.

(7.2.1) Estimation of Suitable Parameter Values for Good Step Response in Inductance-Compensated Amplifier.

The values given in Table 3 (Section 12.4) enable the expression on the right-hand side of eqn. (40) to be calculated. The results for 32 stages are as follows:

K	1.77	1.66	1.55	1.45	1.37
Peak-to-peak ring	0.001	0.06	0.5	2.0	6.9

The results obtained from Reference 12 give

K	1.61	1.51	1.41
Peak-to-peak ring	0.07	0.35	1.15

Comparison of these two sets of results shows that the estimated values are too high by a factor of between two and three times. The approximation made to the integral in eqn. (40) will be in excess for negative exponents, since the peak height will not reach the value inserted to obtain eqn. (40). Also the shape will not be parabolic. However, if a compensation circuit is being investigated, the peak-to-peak ring will be required to be small, and an error of this type is not serious. The value of parameter would not be required accurately, since stray capacitances, for example, could not be known precisely.

As n is altered, the ring would alter slowly, since in this instance the dominant contribution in eqn. (40) comes from the term

$$3 \cdot 3 z_0 (nX_0''')^{1/3}$$

In this case z_0 and F_0 change sign together, so a case of rings of small total amplitude would not be possible.

(7.2.2) Estimation of Location of One Pair of Poles for Good Step Response.

The usual parameter of the circuit of Section 6.2 is the value of the feedback resistance between the anodes. If, however,

this resistance is fixed and the anode loads are varied together, the network function may be reduced to

$$G(z) = \frac{\alpha^2 + 1}{(z + \alpha)^2 + 1} \quad (41)$$

with α as the variable parameter.

The impulse response of n stages of this type is

$$(\alpha^2 + 1)^n e^{-\alpha \tau} \frac{\sqrt{\pi}}{(n-1)!} \left(\frac{\tau}{2} \right)^{n-1/2} J_{n-1/2}(\tau) \quad (42)$$

With this choice of $G(z)$ it may be shown that

$$\begin{aligned} nF_0 + 3 \cdot 3 z_0 \left(\frac{nX_0'''}{2} \right)^{1/3} \\ = n[1 - \alpha + \log(\alpha^2 + 1) - \log 2] + 3 \cdot 3(1 - \alpha) \left(\frac{n}{2} \right)^{1/3} \end{aligned} \quad (43)$$

From the results of Section 7.2.1 it might be anticipated that estimates of ring produced from this equation would be in excess of actual. With 10 stages and $\alpha = 1.3$, eqns. (43) and (40) lead to an estimate of 10%. The calculated response is shown in Fig. 10, and exact evaluation of area between the first and second

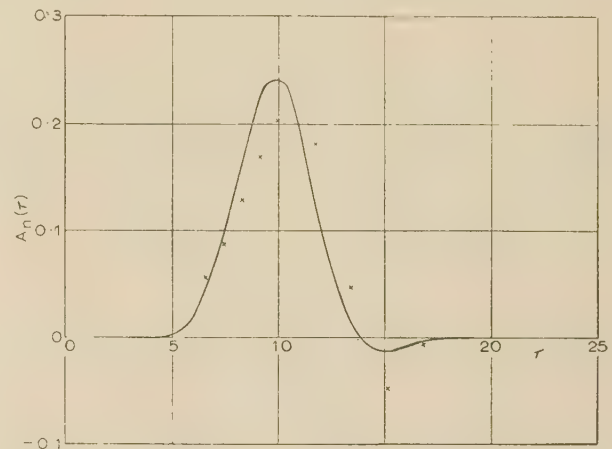


Fig. 10.—Impulse response of 10 stages each of transfer function

$$\frac{2.69}{(z + 1.3)^2 + 1}$$

— Exact.
× Approximate, from eqn. (35).

zeros gives a figure of 2.5%. Values of α of 1.2 and 1.4 give estimates of 16% and 5% respectively. The estimates are once again too high by a few times, but the magnitude of the parameter α obtained would be of practical value.

Also on Fig. 10 is plotted the approximation to the impulse response for $\alpha = 1.3$, given by eqn. (35). It will be seen that the first maximum is about 20% too low and the minimum considerably too high, so that, apart from errors produced by the crude method of integration, the actual approximation is itself not very good. However, if values to within 20% only are required, which is often quite adequate from a practical standpoint, eqn. (35) can be used to give the salient features of delay time, time of rise and ring of the response to a step input.

(8) THE IDEAL LOW-PASS AMPLIFIER

A cascade of the Wheeler circuits of Section 6.1 approximates to the case of constant gain in the pass band with great attenuation outside. In the limit, it becomes the ideal case with zero gain outside the pass band. It is therefore of interest to see what the above approximations give for this case. Since a

large number of circuits is to be considered, eqn. (35) will serve as an adequate approximation. In this case $z_0 = 0$, $\lambda_0 = 1$, $X_0''' = 1$, $F_0 = 0$. Hence eqn. (35) gives

$$A_n(\tau) \simeq \left(\frac{2}{n}\right)^{1/3} \text{Ai}\left[-\left(\frac{2}{n}\right)^{1/3}(\tau - n)\right] \quad (44)$$

The maximum of the Airy function, from Fig. 8, is about 0.54. Hence the maximum slope of the step response decreases as $n^{-1/3}$.

The amplitude difference between first maximum and first minimum is

$$\int \text{Ai}(-x)dx$$

between its first and second zeros. The response to a step input is therefore of the usual type, with delay proportional to n , a maximum slope decreasing as $n^{-1/3}$ and a ring of constant amplitude.

(9) CONCLUSIONS

Analytical expressions have been obtained which are approximations to the impulse response of many stages in cascade. The expressions given in Section 5 have been shown to give useful results for only a small number of stages in two particular cases. The simpler but less accurate expressions in Section 7 have been shown to be applicable to certain types of problem. These apply only to the particular pole-zero distributions which have the critical value of λ determined by $X'' = 0$. A similar approximation for more complicated cases can probably be obtained.

Apart from their possible use in obtaining numerical answers, the approximations may be useful in a semi-quantitative description of the transient response of circuits.

(10) ACKNOWLEDGMENTS

Thanks are due to Miss G. Smith for assistance in computing the results quoted in Section 7, and to Mr. J. G. Yates for helpful discussions.

(11) REFERENCES

- (1) ROSENBROCK, H. H.: 'An Approximate Method for Obtaining Transient Response from Frequency Response', (*Proceedings I.E.E.*, Paper No. 1907 M, November, 1955 (102, Part B, p. 744).
- (2) CARSON, J. R.: 'The Building-up of Sinusoidal Currents in Long Periodically Loaded Lines,' *Bell System Technical Journal*, 1924, 3, p. 558.
- (3) EAGLESFIELD, C. C.: 'Carrier-frequency Amplifiers', *Wireless Engineer*, 1945, 22, p. 523.
- (4) WATSON, G. N.: 'A Treatise on the Theory of Bessel Functions', 2nd Edition (Cambridge, 1944), p. 229.
- (5) MORSE, P. M., and FESHBACH, H.: 'Methods of Theoretical Physics' (McGraw Hill, 1953), Pt. I, Ch. 4, Section 4.6.
- (6) JEFFREYS, H., and JEFFREYS, B. S.: 'Methods of Mathematical Physics' (Cambridge, 1946), p. 472.
- (7) WATSON, G. N.: Reference 4, p. 235.
- (8) DARLINGTON, S.: 'The Potential Analogue Method of Network Synthesis', *Bell System Technical Journal*, 1951, 30, p. 315.
- (9) BOOTHROYD, A. R., CHERRY, E. C., and MAKAR, R.: 'An Electrolytic Tank for the Measurement of Steady-State Response, Transient Response, and Allied Properties of Networks', *Proceedings I.E.E.*, Paper No. 835, May, 1949 (96, Pt. I, p. 163).
- (10) WHEELER, H. A.: 'Wide-band Amplifiers for Television', *Proceedings of the Institute of Radio Engineers*, 1939, 27, p. 429.

(11) WATSON, G. N.: Reference 4, p. 516.

(12) BEDFORD, A. V., and FREDENDALL, G. L.: 'Transient Response of Multistage Video-Frequency Amplifiers', *Proceedings of the Institute of Radio Engineers*, 1939, 27, p. 277.

(13) WHITTAKER, E. T., and WATSON, G. N.: 'A Course of Modern Analysis', 4th edition (Cambridge, 1952), Ch. XII, Section 12.33.

(14) Tables of Modified Hankel Functions of Order One-third and their Derivatives' (Harvard University Press, 1945).

(12) APPENDICES

(12.1) Method of Calculation to obtain Results given in Section 2.4

To determine z , σ is expanded as a power series about the point z_c and the series is then inverted. This can be done by evaluating the derivatives of σ at z_c or in this case by expanding the logarithm. Put $z + 1 - \frac{1}{\lambda} = \zeta$; then

$$\begin{aligned} \sigma &= 1 - \lambda\left(\zeta + \frac{1}{\lambda}\right) + \log(\lambda\zeta + 1) \\ &= -\frac{\lambda^2\zeta^2}{2} + \frac{\lambda^3\zeta^3}{3} - \frac{\lambda^4\zeta^4}{4} + \dots \end{aligned}$$

from which $\lambda\zeta = \pm j(2\sigma)^{1/2} + a\sigma + b\sigma^{3/2} + \dots$

and by substitution it may be shown that

$$\lambda\zeta = \pm j(2\sigma)^{1/2} - \frac{\sqrt{2}}{3}\sigma \mp j\frac{\sqrt{2}}{18}\sigma^{3/2} + \dots$$

The two values of z are given by the plus and minus signs. Let z_2 have the plus sign. Then

$$z_2 - z_1 = 2j\mathcal{J}(\zeta_2) = \frac{2j\sqrt{2}}{\lambda}\left(\sigma^{1/2} - \frac{\sigma^{3/2}}{18} + \dots\right)$$

and $\frac{d}{d\sigma}(z_2 - z_1) = \frac{j\sqrt{2}}{\lambda}\left(\sigma^{-1/2} - \frac{1}{6}\sigma^{1/2} + \dots\right)$

We therefore have

$$RCG_0^{-n}A_n(\tau) = \frac{\varepsilon^{nF_c}}{2\pi j} \int_0^\infty \varepsilon^{-n\sigma} \frac{j\sqrt{2}}{\lambda} \left(\sigma^{-1/2} - \frac{1}{6}\sigma^{1/2} + \dots\right) d\sigma$$

It can be shown that the series may be integrated term by term to yield an asymptotic series. Hence

$$\begin{aligned} RCG_0^{-n}A_n(\tau) &\sim \frac{\varepsilon^{nF_c}}{\pi} \frac{1}{\lambda\sqrt{2}} \left[\Gamma(1/2)n^{-1/2} - \frac{1}{6}\Gamma(3/2)n^{-3/2} + \dots \right] \\ &\sim \frac{\varepsilon^{nF_c}}{\pi} \frac{1}{\sqrt{(2\pi n)}} \left(1 - \frac{1}{12n} + \dots\right) \end{aligned}$$

Substituting for F_c ,

$$\begin{aligned} RCG_0^{-n}A_n(\tau) &\sim \frac{1}{\lambda} \frac{\varepsilon^{n(1-\lambda+\log\lambda)}}{\sqrt{(2\pi n)}} \left(1 - \frac{1}{12n} + \dots\right) \\ &\sim \frac{\varepsilon^{n-\tau}}{\sqrt{(2\pi)}} \tau^{n-1} n^{-n+1/2} \left(1 - \frac{1}{12n} + \dots\right) \end{aligned}$$

which is eqn. (13).

Stirling's asymptotic expression¹³ for $\Gamma(x)$ is

$$\Gamma(x) \sim \varepsilon^{-x} x^{x-1/2} \sqrt{(2\pi)} \left(1 + \frac{1}{12x} + \dots\right)$$

Hence

$$\frac{1}{(n-1)!} = \frac{1}{\Gamma(n)} \sim \frac{\varepsilon^n}{\sqrt{(2\pi)} n^{-n+1/2}} \left(1 - \frac{1}{12n} + \dots\right)$$

By using this in eqn. (6), eqn. (13) is obtained.

(12.2) Method of Calculation to obtain Results given in Section 5.1

The contour C may be determined directly by sketching the curve satisfying the equation

$$\mathcal{J}(\sigma) = 0$$

For α_2 greater and less than zero it is found that

$$J_1 = \int_C \varepsilon^{-n\bar{\sigma}} dz = \int_{\infty e^{-j\pi/3}}^{\infty e^{j\pi/3}} \exp(n\alpha_1 \zeta^2 + n|\alpha_2| \zeta^3) d\zeta \quad (45)$$

If the variable is changed to w , defined by

$$\zeta = \frac{w}{(3n|\alpha_2|)^{1/3}} - \frac{\alpha_1}{3|\alpha_2|} \quad (46)$$

it is found that

$$n\alpha_1 \zeta^2 + n|\alpha_2| \zeta^3 = \frac{w^3}{3} + \frac{2\mu^{3/2}}{3} - \mu w \quad (47)$$

where

$$\mu = \frac{n^{2/3} \alpha_1^2}{(3|\alpha_2|)^{4/3}} \quad (48)$$

Hence

$$J_1 = \frac{\exp(\frac{2}{3}\mu^{3/2})}{(3n|\alpha_2|)^{1/3}} \int_{\infty e^{-j\pi/3}}^{\infty e^{j\pi/3}} \varepsilon^{w^3/3 - \mu w} dw \quad (49)$$

This integral can be evaluated in terms of modified Hankel functions or Airy functions. The two relevant definitions are¹⁴

$$h_1(z) = \frac{k}{j\pi} \int_{-\infty}^{\infty e^{j\pi/3}} \varepsilon^{zt + t^3/3} dt \quad (50)$$

$$h_2(z) = \frac{k^*}{j\pi} \int_{-\infty}^{\infty e^{-j\pi/3}} \varepsilon^{zt + t^3/3} dt \quad (51)$$

in which

$$k = 12^{1/6} \varepsilon^{-j\pi/6}$$

The Airy function $\text{Ai}(z)$ is connected with these by the relation

$$\text{Ai}(z) = \frac{1}{2k} h_1(-z) + \frac{1}{2k^*} h_2(-z) \quad (52)$$

Hence eqn. (49) becomes

$$\begin{aligned} J_1 &= \frac{\exp(\frac{2}{3}\mu^{3/2})}{(3n|\alpha_2|)^{1/3}} \left[\frac{j\pi}{k} h_1(-\mu) - \frac{-j\pi}{k} h_2(-\mu) \right] \\ &= 2\pi j \frac{\exp(\frac{2}{3}\mu^{3/2})}{(3n|\alpha_2|)^{1/3}} \text{Ai}(\mu) \quad (53) \end{aligned}$$

It may be observed that since α_2 changes sign between the critical point and infinity on the real axis, it must vanish at some point, and hence μ must become infinite at this point. Using the asymptotic expansion¹⁴ for $\text{Ai}(\mu)$ it may be shown that at the point for which $\alpha_2 = 0$

$$J_1 = j \left(\frac{\pi}{n\alpha_1} \right)^{1/2} \quad (54)$$

(12.3) Method of Calculation to obtain Results given in Section 5.2

It is required to evaluate the integral

$$J_2 = \int \exp(n\alpha_1 \zeta^2 + n\alpha_2 \zeta^3) d\zeta \quad (55)$$

over a contour for which the exponent is real and negative. Since α_1 and α_2 are now complex the determination of the appropriate contour is more difficult than in Section 12.2.

$$-\bar{\sigma} = n\alpha_1 \zeta^2 + n\alpha_2 \zeta^3 \quad (56)$$

First, change the variable to t defined by

$$\frac{t^3}{3} = n\alpha_2 \zeta^3 \quad (57)$$

This corresponds to a rotation of the axes in the ζ -plane. It is then found that

$$-\bar{\sigma} = \frac{n\alpha_1}{(3\alpha_2 n)^{2/3}} t^2 + \frac{t^3}{3} \quad (58)$$

$$= \mu^{1/2} t^2 + \frac{t^3}{3} \quad (59)$$

A translation of axes by change to a variable w , defined by

$$w = t + \mu^{1/2} \quad (60)$$

gives

$$-\bar{\sigma} = \frac{2}{3}\mu^{3/2} - \mu w + w^3/3 \quad (61)$$

To determine the contours it is easiest to take eqn. (59), since the col is at the origin in the t -plane.

Investigation of the equation $\mathcal{J}(\bar{\sigma}) = 0$ leads to the series of contours in the t -plane shown diagrammatically in Fig. 11, in

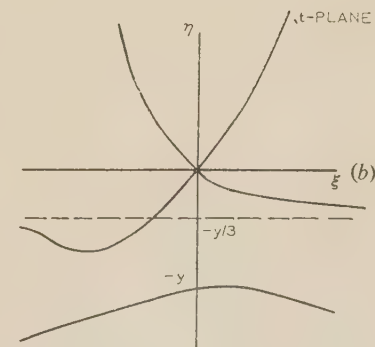
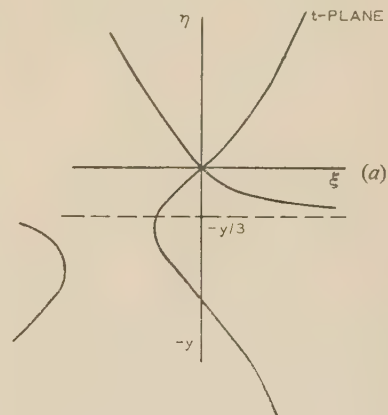


Fig. 11.—Contours of $\mathcal{J} \left[(x + jy)t^2 + \frac{t^3}{3} \right] = 0$.

(a) $3x^2 > y^2$; $x, y > 0$.

(b) $3x^2 < y^2$; $x, y > 0$.

For $x < 0$ diagrams are reversed about $\xi = 0$.
For $y < 0$ diagrams are reversed about $\eta = 0$.

which $\mu^{1/2} = x + jy$. For x negative the diagrams are reversed about the η axis, and for y negative about the ξ axis. All the contours other than those asymptotic to $\eta = -y/3$ are, for large values of ξ and η , straight lines of slope $\pm\sqrt{3}$. Consideration of the expression for $\mathcal{R}(\sigma) \simeq \mathcal{R}(t^{3/3})$ for large values of $|t|$ shows that this can only be negative for points whose arguments are within the ranges $\pi/6$ to $\pi/2$, $-\pi/6$ to $-\pi/2$, $5\pi/6$ to $7\pi/6$.

It is simple to determine the values of x and y for λ near λ_0 by expanding the various quantities in a series of powers of $(\lambda - \lambda_0)^{1/2}$. It is found that

$$\mu^{1/2} \simeq jn^{1/3} \left(\frac{\chi_0'''}{2} \right)^{-1/6} (\lambda - \lambda_0)^{1/2} \quad (62)$$

In the neighbourhood of the critical point, therefore, $\mu^{1/2}$ is positive imaginary. Hence, referring to Fig. 11(b), $3x^2 < y^2$, $y > 0$, and whatever the sign of x the appropriate contour is $(-\infty, \infty e^{j\pi/3})$. This is a complete designation, since there are no singularities in the finite part of the plane. This is also the contour in the w -plane. It will be assumed that this is the contour valid for values of λ giving cols away from the critical point, unless detailed analysis in particular cases shows otherwise.

Eqn. (55) therefore becomes

$$\begin{aligned} J_2 &= \frac{1}{(3n\alpha_2)^{1/3}} \int_{-\infty}^{\infty e^{j\pi/3}} \exp \left(\frac{2}{3} \mu^{3/2} - \mu w + \frac{w^3}{3} \right) dw \\ &= \frac{\exp(\frac{2}{3} \mu^{3/2})}{(3n\alpha_2)^{1/3}} \frac{j\pi}{k} h_1(-\mu) \quad (63) \end{aligned}$$

(12.4) The Inductance-Compensated Video-Amplifier

The schematic circuit is shown in Fig. 9. The gain is given by

$$g \left(\frac{Lp + R}{1 + RCp + LCP^2} \right) \quad (64)$$

If reduced variables

$$\tau = \frac{Rt}{L}, z = \frac{Lp}{R}, m = \frac{L}{R^2C} \quad (65)$$

are used, and the gain is normalized to be unity at low frequencies, it is found that

$$\chi = \log m(1+z) - \log(m+z+z^2) \quad (66)$$

From this expression the first three derivatives may be determined, and by equating the expression for the second derivative to zero, an equation may be set up between z and m for determining the critical point. The nature of the equation obtained makes it easier to specify the critical point z_0 and then find the value of m which satisfies the equation. Two sets of results have been computed, one for two specific values of m used by Bedford and Fredendall,¹² and the other for a range of positions of the critical point. These are given Tables 2 and 3. The parameter used in Reference 12 is $K = m^{-1/2}$, and the time variable $t/2\pi RC = m\tau/2\pi$.

Table 2

$K = m^{-1/2}$	m	z_0	λ_0	χ_0'''
1.41	0.5000	1.1484	1.064	4.158
1.61	0.3864	-0.0543	1.606	12.28

Table 3

z_0	λ_0	χ_0	χ_0'''	F_0	m	K
-0.2	2.49	0.469	39.2	-0.029	0.321	1.77
-0.1	1.81	0.180	16.9	-0.001	0.364	1.66
0.0	1.42	0	8.8	0	0.416	1.55
0.1	1.14	-0.113	4.12	0.001	0.474	1.45
0.2	0.98	-0.191	3.32	0.005	0.532	1.37

THE USE OF LOSSY MATERIAL TO SUPPRESS UNWANTED MODES IN CAVITY RESONATORS

By M. Y. EL-IBIARY, B.Sc., Ph.D., Graduate, and J. BROWN, M.A., Ph.D., Associate Member.

The paper was first received 27th February and in revised form 27th April, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

Unwanted modes in cylindrical cavity resonators operating in the H_{01n} mode may be suppressed by inserting a ring of lossy material in a groove cut at the junction between an end wall and the cylindrical wall. This groove does not affect the H_{01n} mode but may suppress the others if the groove dimensions are correctly chosen. An analysis of the effectiveness of this method is given and is confirmed by measurements in the 3 cm band of wavelengths.

LIST OF SYMBOLS

- a = Radius of cavity.
 r = Inner radius of groove.
 $w = a - r$ = width of groove.
 c = Length of cavity.
 d = Depth of groove.
 λ_0 = Wavelength
 β_0 = Phase-change coefficient
 Y_0 = Wave admittance
 k = Turns ratio of ideal transformer
 B = Susceptance
 ρ, ϕ, z = Cylindrical co-ordinates.
 β_{lm} = Phase-change coefficient
 Y_{lm} = Wave admittance
 $E_m(\rho, \phi)$ = Transverse electric field
 A_m = Amplitude
 X_{lm} = m th root (excluding zero) of $J_l'(x) = 0$.
 X'_{lm} = m th root (excluding zero) of $J_l(x) = 0$.
 $\gamma_g = \alpha_g + j\beta_g$, propagation coefficient
 α_g = Attenuation coefficient
 β_g = Phase-change coefficient
 Y_g = Wave admittance
 C = Amplitude
 ϵ_r, μ_r = Relative permittivity and permeability of material filling the groove.
 u = Decay factor.
 Q_0 = Unloaded Q-factor of cylindrical cavity.
 Q_{rad} = Q-factor resulting from losses in the coupling holes.
 Q_g = Q-factor resulting from losses in the groove.
 \bar{Q} = Q-factor of cavity-groove combination.
 $\delta(x)$ = Dirac delta function.
 ν_l = Neumann's number; $\nu_0 = 1 : \nu_l = 2$ if $l \neq 0$.
 $b_0 = B/k^2 Y_g$; $K = \beta_g^2 c k^2 Y_g / \beta_{lm} Y_{lm}$
 $F_{lm} = (X_{lm}^2 - l^2) / \nu_l l^2$.
 f_m, f'_m = Parameters defined by eqns. (40) and (41).
 S_l, S'_l = Series defined by eqns. (62) and (63).

The M.K.S. system of units is used.

(1) INTRODUCTION

A cavity resonator has an infinite number of modes of oscillation, each one typified by its resonant frequency and field

distribution. In general, the resonant frequencies are quite separate, but occasionally two different modes may occur at the same frequency or at two very nearly equal frequencies. In such a situation the properties of the modes, in particular their Q-factors, may be quite different from those when the modes are widely separated in frequency. Certain applications of cavities, e.g. echo boxes, require a mode with as high a Q-factor as possible, and it has been shown by Kinzer and Wilson¹ that the most suitable cavity for such a purpose is a circular cylinder operating in an H_{01n} mode. The range of frequency for which such a resonator may be used is restricted by the occurrence of other modes at the same frequency as the H_{01n} when the cavity is tuned. Methods for the elimination of these other unwanted modes are therefore of considerable practical importance because they permit the possible frequency range of the H_{01n} cavity to be extended.

Three methods are known for the suppression of unwanted modes, namely:

- (a) selection of the cavity dimensions,
- (b) design of the coupling between the external waveguide and the cavity to avoid the excitation of these modes, and
- (c) insertion of lossy material at a position in the cavity where it will damp out the unwanted modes without materially affecting the wanted mode.

The first of these has been fully described by Wilson *et al.*² and is commonly used by cavity designers. An example of the second method has been described by Bleaney *et al.*,³ and leads to a considerable increase in the usable frequency range. In principle, more complicated coupling systems suppressing many modes can be designed but lead to constructional problems. Further, although the external guide may not couple to an unwanted mode, departures from the ideal cylindrical shape may lead to cross-coupling within the cavity between this mode and the desired one, leading to a drop in the Q-factor of the desired mode.^{4,5} The third possibility has been suggested by several workers, and a commonly used example is the elimination of the E_{11n} mode by lossy material on the rear surface of the tuning plunger of an H_{01n} resonator. In the present paper, a similar method is investigated theoretically and experimentally to determine what degree of suppression can be achieved.

(2) USE OF A LOSSY GROOVE IN THE CAVITY END WALL

(2.1) Resonance Condition

The lossy material must be introduced into the cavity in such a position that it does not affect the wanted mode. A convenient method of doing this is to cut a groove along a line of current flow for this mode on the cavity wall and to fill it with the lossy material. The current for the wanted mode does not flow across the groove and so is only slightly affected by its presence provided that its width is not too great. Any modes which have lines of current flow crossing the groove are considerably modified and will be suppressed to an extent depending on the groove dimensions and the nature of the filling.

Correspondence on Monographs is invited for consideration with a view to publication.
 Dr. El-Ibiary is in the Electrical Engineering Department, Cairo University.
 Dr. Brown is in the Electrical Engineering Department, University College, London.
 The authors were in the Electrical Engineering Department, Imperial College of Science and Technology, University of London, when the investigation described was carried out.

This technique has been investigated for an H_{01n} resonator, the position selected for the groove being at the junction between the end wall and the cylindrical wall, as in Fig. 1. This groove

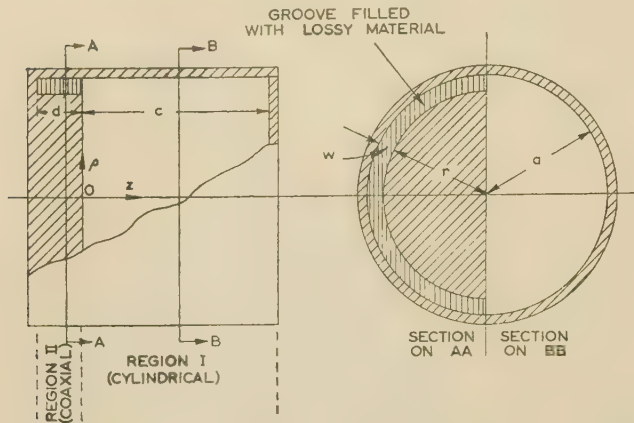


Fig. 1.—Sketch of the groove and cavity.

intersects the lines of current flow for all the cavity modes except those of H_{0mn} type. The performance of the groove can be most easily discussed with the help of an equivalent circuit, which can be deduced from the nature of the fields in the modified cavity. The cavity-groove combination can be regarded as a junction between a circular waveguide and a coaxial waveguide. If the frequency is near the resonant frequency of the H_{lmn} mode for the perfect cylinder, the most important contributions to the fields in region I, the circular waveguide, arise from a pair of H_{lm} circular waveguide modes travelling in opposite directions. These waves can be represented by a transmission line, L_1 , having the same phase-change coefficient, which is short-circuited at $z = c$ corresponding to the end wall of the cavity (Fig. 2). Since

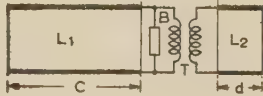


Fig. 2.—Equivalent circuit for the cavity-groove combination.

L_1 is a transmission line representing the principal waveguide mode in the cylinder, L_2 is a transmission line representing the coaxial mode in the groove, T is a coupling transformer of turns ratio k , B is a susceptance corresponding to evanescent modes in the cavity and the groove.

the cavity is axially symmetrical, all fields, both in region I and in region II, must have the same angular periodicity as the H_{lmn} mode. For narrow grooves, the only possible propagating mode is then the H_{l1} coaxial mode, and this can also be represented by a length of transmission line, L_2 , having the propagation coefficient of the H_{l1} mode and being short-circuited at $z = -d$, d being the depth of the groove. Since the groove is usually filled with lossy material, the propagation coefficients γ_g and the wave admittance Y_g for this mode are in general both complex. The imaginary part of Y_g is sufficiently small to be neglected, and in the remainder of the paper Y_g will be assumed to be purely real.

The two modes are coupled together at the junction plane, $z = 0$, and in the equivalent circuit this coupling takes the form of an ideal transformer of turns ratio k , provided that both transmission lines have characteristic admittances equal to the wave admittances of the corresponding waveguide modes. Other modes will be excited in both the cavity and the groove, but with exceptions to be discussed later these modes will be evanescent. They can be lumped together and represented in the circuit by the susceptance B . The parameters k and B can

be calculated from the field distributions in the cavity and the groove.

The condition that the circuit of Fig. 2 be resonant is:

$$jY_{lm} \cot(\beta_{lm}c) = jB + k^2 Y_g \coth(\gamma_g d) \quad (1)$$

in which the only unknown is the frequency. The quantities β_{lm} , β_g , Y_{lm} , Y_g , B and k are all functions of frequency, which makes it virtually impossible to solve eqn. (1) exactly. In the case in question, however, the resonant frequency of the modified cavity will differ only slightly from that of the perfect cylindrical cavity, and the equation can then be used to determine the shift in the resonant frequency caused by the groove. It is more convenient to work with wavelengths, and if λ_0 is the free-space wavelength for resonance of the H_{lmn} mode in the ideal cylindrical cavity,

$$(2\pi/\lambda_0)^2 = (\beta_{lm}^0)^2 + (X_{lm}/a)^2 \quad (2)$$

where

$$\beta_{lm}^0 c = n\pi \quad (3)$$

and X_{lm} is the m th root other than zero of $J'_l(x) = 0$.

Suppose now that the effect of the groove is to change the resonant wavelength to $\lambda_0 + \delta\lambda$, where $\delta\lambda$ is much less than λ_0 . The term $\cot(\beta_{lm}c)$, which is infinite for the value of β_{lm} given by eqn. (3), will change very rapidly with $\delta\lambda$ and to a good approximation is given by

$$\cot(\beta_{lm}c) = 1/c\delta\beta \quad (4)$$

where $\delta\beta$ is the change in β_{lm} corresponding to the change $\delta\lambda$ in the resonant wavelength. Now from eqn. (2),

$$\beta_{lm}\delta\beta = -\left(\frac{2\pi}{\lambda_0}\right)^2 \frac{\delta\lambda}{\lambda_0} \quad (5)$$

The other wavelength-dependent terms in eqn. (1) will change very much less near λ_0 and can be evaluated at λ_0 without significant error. The shift $\delta\lambda$ is therefore given by

$$-j \frac{Y_{lm}\beta_{lm}}{\beta_{lm}^0 c} \frac{\lambda_0}{\delta\lambda} = jB + k^2 Y_g \coth(\gamma_g d) \quad (6)$$

where β_0 has been written for $2\pi/\lambda_0$.

The wave admittances are given by

$$Y_{lm} = \beta_{lm} Y_0 / \beta_0 \quad (7)$$

$$Y_g = \beta_g Y_0 / \mu_r \beta_0 \quad (8)$$

Y_0 being the admittance of a plane wave in free space, and μ_r the relative permeability of the material in the groove. Eqn. (8) is correct to the order of approximation already mentioned, i.e. neglecting the imaginary part of Y_g .

$$\frac{j(\beta_{lm}/\beta_0)^2 \lambda_0}{c} \frac{1}{\delta\lambda} = -\frac{k^2}{\mu_r} \beta_g \coth(\gamma_g d) - j \frac{\beta_0 B}{Y_0} \quad (9)$$

Eqn. (9) has been derived in this Section by physical arguments and the quantities k and B are as yet undetermined. A more rigorous analysis is given in Section 7.1 and leads to a similar equation (44) justifying the present argument. Further, comparison with this equation gives values for k and B .

(2.2) The Effect of the Groove on the Q-factor

When the groove is filled with a lossy material, the right-hand side of eqn. (9) is complex, so that the calculated value of $\delta\lambda$ has an imaginary part, i.e. free oscillations in the cavity can only occur for complex frequencies. The time-dependence of the field is then of the type $\exp(j\omega t - ut)$, u , the imaginary part of the frequency, being interpreted as a decay factor. This decay factor is related to the Q-factor by the equation

$$Q_g = \omega/2u \quad (10)$$

where the symbol Q_g has been used to emphasize that this Q-factor is solely due to the losses in the groove. There will be other losses due to the wall conductivity of the main cavity and the coupling to the waveguides and these will cause the Q-factor of the mode being considered to be Q_0 , say, when the groove is absent. From the usual result for combining Q-factors, it follows that the effective Q-factor for the cavity-groove combination is Q , given by

$$1/Q = 1/Q_0 + 1/Q_g \quad (11)$$

The value of Q_g may be calculated from eqn. (9), but before this is done it will be helpful to discuss quantitatively the results to be expected. The assumption that the right-hand side of eqn. (1) can be evaluated at the resonant frequency for the ideal cylindrical cavity will again be made. The feature of major interest then is the variation of the Q-factor due to the groove as the depth d of the filling is varied. The term $k^2 Y_g \coth(\gamma_g d)$ is the admittance of the groove transformed to the right of the ideal transformer in the equivalent circuit of Fig. 2 and can be written

$$k^2 Y_g \coth(\gamma_g d) = k^2 Y_g (g + jb) \quad (12)$$

where g and b are respectively the normalized conductance and susceptance of the short-circuited length of line corresponding to the groove. If this expression is substituted in eqn. (6), there is obtained after a little rearrangement

$$\frac{\lambda_0}{\delta \lambda} = \frac{[-B - k^2 Y_g b + jk^2 Y_g g] \beta_0^2 c}{\beta_{lm} Y_{lm}} \quad (13)$$

The corresponding frequency shift is $\delta \omega$, where

$$\frac{\omega}{\delta \omega} = -\frac{\lambda_0}{\delta \lambda} = \frac{[B + k^2 Y_g b - jk^2 Y_g g] \beta_0^2 c}{\beta_{lm} Y_{lm}} \quad (14)$$

and if $\delta \omega$ is written as $\delta + j\epsilon$, δ, ϵ both being real, the imaginary part is found to be

$$\epsilon = \frac{\omega_0 g}{K[(b + b_0)^2 + g^2]} \quad (15)$$

where

$$b_0 = B/k^2 Y_g \quad (16)$$

and

$$K = \beta_0^2 c k^2 Y_g / \beta_{lm} Y_{lm} \quad (17)$$

Since g must always be positive, ϵ is positive, confirming that the resonant frequency for the cavity-groove combination has a positive imaginary part. The quantity u in eqn. (10) is equal to ϵ , and if δ is assumed to be much less than ω , the Q-factor becomes

$$Q_g = \frac{K[(b + b_0)^2 + g^2]}{2g} \quad (18)$$

The only terms in this equation which depend on the groove depth d are b and g , and the problem now is to determine what value d makes Q_g a minimum, this corresponding to the maximum degree of suppression that can be achieved. The terms b and g vary with d in a fairly complicated manner, and the answer is most easily obtained by using a Smith chart to show the variation. This is done in Fig. 3, where the curve OALM . . . shows the locus of the admittance, starting from the point 0, corresponding to the short-circuit at $d = 0$. Eqn. (18) shows that Q_g can never be less than $Kg/2$, so that the smallest possible value of Q_g would arise for the point A, provided that b_0 were zero. If, however, b_0 is not zero, the term $K(b + b_0)^2/g$ exerts an influence, and since b is changing much more rapidly near L than is g , it appears that a very good approximation to the minimum value of Q_g is obtained at the point L, for which b is equal to $-b_0$. As d is increased further, a set of minima will occur for the series of points M, N, . . . , which are spaced at approximately half-

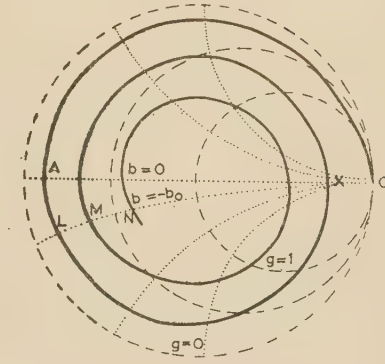


Fig. 3.—Locus of the normalized admittance of the groove as the depth varies, sketched on a Smith chart.

wavelength intervals in d . The diagram shows, however, that the values of g at these minima become larger as d increases. The conclusion is therefore that the greatest suppression of the mode will occur for the point L, i.e. when d has the smallest value that makes b equal to $-b_0$ with g less than unity. This restriction excludes point X, for which b is also equal to $-b_0$.

The form of eqn. (18) suggests that complete suppression of the mode could be achieved by arranging $(b + b_0)$ and g both to be zero, which could in principle be done by a loss-free groove. This is, of course, at variance with physical ideas, and the reason for the discrepancy lies in the assumption that the resonant frequency of the cavity-groove combination differs only slightly from that of the ideal cavity. Reference to eqn. (14) shows that when $(b + b_0)$ and g both tend to zero, $\delta \omega$ tends to infinity, so that the assumptions on which eqn. (18) is based are no longer valid. Further, when the real part of $\delta \omega$ becomes appreciable it may be necessary to allow for the frequency variation of both the terms B and $Y_g \coth(\gamma_g d)$. It is evident from the general circuit shown in Fig. 2 that the cavity-groove combination can be regarded as a pair of coupled circuits, and if d approaches the value to make the groove resonant at nearly the same frequency as the cavity itself, two resonances will occur. This is shown analytically by the frequency shift becoming double-valued; experimental confirmation is discussed in Section 3.2. In practical applications of the method the losses of the material in the groove will suffice to prevent double resonances occurring.

The optimum length for d is given approximately by the condition that $(b + b_0)$ should be zero, so that, if β_g is the real part of γ_g ,

$$B + k^2 Y_g \cot(\beta_g d) = 0 \quad (19)$$

provided that the attenuation given by the imaginary part of γ_g is much less than β_g . When b_0 is zero, d is a quarter of the wavelength of the mode in the groove (corresponding to point A in Fig. 3) and d is less than or greater than a quarter-wavelength when b_0 is negative or positive respectively. If the groove is designed for the greatest mode suppression, d will never exceed half the wavelength.

When the depth is adjusted to its optimum point, the value of Q_g is $Kg/2$ and is therefore dependent on K . From eqn. (17) and the results in Section 7.2, K may be written

$$K = \frac{ac \beta_0^2 Y_g F_{lm}}{\omega \beta_{lm} Y_{lm}} \quad (20)$$

F_{lm} being tabulated in Table 2. The factors which vary appreciably with mode type are F_{lm} , β_{lm} and Y_{lm} , the latter being proportional to β_{lm} , so that

$$K \propto F_{lm} / \beta_{lm}^2 \quad (21)$$

At resonance, eqn. (3) is satisfied, giving

$$K \propto F_{lm}/n^2 \quad \dots \quad (22)$$

The modes which are most easily suppressed are those with small values of K , and it follows from eqn. (22) and Table 2 that modes with large values of l and n and small values of m are the most easily dealt with.

(3) EXPERIMENTAL RESULTS

(3.1) The Design of a Cavity

Experiments have been carried out in the 3 cm band of wavelengths to confirm the conclusions of Section 2. For simplicity, it was decided to use a cavity of fixed dimensions, since in a tunable cavity the gap between the plunger and the cylindrical wall would have an unknown effect on the higher-order modes whose suppression it was desired to investigate. The principal considerations affecting the choice of dimensions were that as many modes as possible were desired in the frequency range available from the oscillator, but that these modes should be reasonably well separated in frequency. Inspection of the cavity mode chart² showed that a length of 7.11 cm and a diameter of 5.64 cm were suitable; these give seven H modes and five E modes within the wavelength range 3.01–3.34 cm. The H modes are H_{013} , H_{114} , H_{121} , H_{214} , H_{313} , H_{411} and H_{412} . Two diametrically opposite $\frac{1}{4}$ in holes, drilled in the cylindrical wall 3 cm from one end plate, couple to rectangular waveguides positioned with their longer sides parallel to the cavity axis. The waveguides are recessed into the cavity walls to keep the thickness of the wall containing the coupling holes reasonably small. This arrangement couples the waveguides to any of the H modes but avoids direct excitation of the E modes.

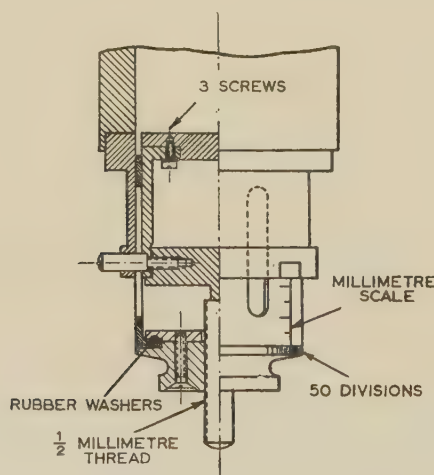


Fig. 4.—Arrangement of end-plate, permitting the depth of the groove to be varied.

A special end-plate was constructed as shown in Fig. 4 having a groove of width 2 mm positioned to give the arrangement already shown in Fig. 1. A circular liner can be moved within this groove by a micrometer head to enable any required depth of groove to be obtained.

(3.2) Experiments on the Empty Groove

When the groove is empty, losses can be neglected and the only effect of the groove is then to change the resonant wavelength. The shift in resonant wavelength was plotted as a function of the groove depth; a typical curve, that for the

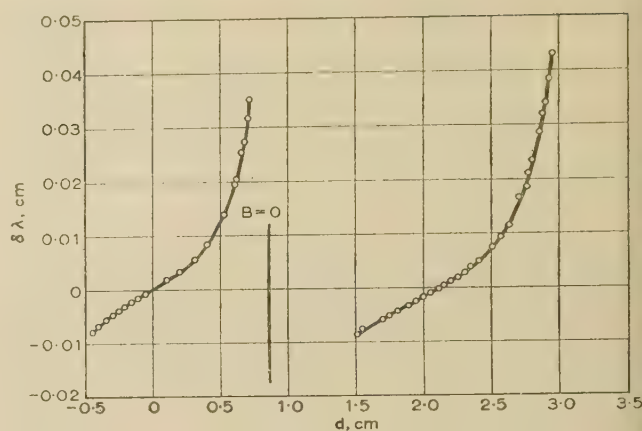


Fig. 5.—Change in the resonant wavelength of the H_{412} mode caused by an empty groove.

H_{412} mode, is shown in Fig. 5. The other modes, excluding the H_{013} , showed similar behaviour except that there were quite marked differences in the magnitudes of the frequency shift. These differences arise from changes in the transformer turns ratio (Fig. 2), which is a measure of the effectiveness of the coupling between the main cavity and the groove. The curve for the H_{013} mode in Fig. 6 shows that changes in the groove depth do not affect the resonant frequency, confirming that the presence of the groove does not materially change the operating conditions for this mode. In Figs. 5 and 6 negative values of $\Delta\lambda$ correspond to the groove plunger extending into the main cavity.

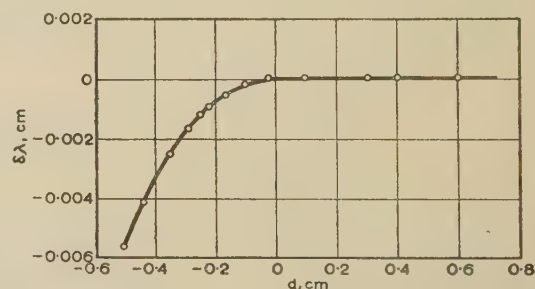


Fig. 6.—Change in the resonant wavelength of the H_{013} mode caused by an empty groove.

For all modes except the H_{013} there is a very rapid change of resonant frequency near a critical depth and this is accompanied by a reduction in the strength of the signal transmitted through the cavity. At this critical depth a second resonance is detected at a shorter wavelength, and the strength of this resonance increases with further increase in the groove depth. From the argument given in Section 2.2 this critical depth is the one which makes the total susceptance, $b + b_0$, zero; in principle, therefore the susceptance, b_0 , can be deduced by inserting in eqn. (1) the experimentally determined values of d and β_g , read from curves such as that in Fig. 5. A difficulty arises in that for some of the modes two resonances exist over a finite range of values of the groove depth, so that no unique value for the critical depth can be obtained. A procedure which leads to a fuller understanding of why this should happen, and which also gives estimates for the transformer turns ratio, is to plot $1/\Delta\lambda$ against $\beta_g \cot(\beta_g d)$. Eqn. (9) shows that this should give a straight line with slope $k^2 c \beta_0 / \lambda_0 \beta_{lm}^2$ (γ_g being equal to $j\beta_g$ and ϵ_r being unity for the empty groove) and intersecting the $\beta_g \cot(\beta_g d)$ axis at the value

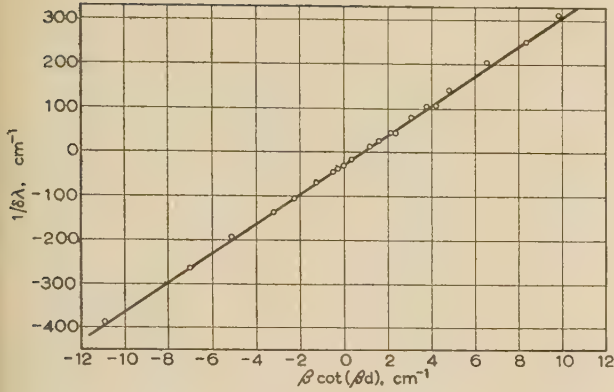


Fig. 7.—Curve of $1/\delta\lambda$ against $\beta_g \cot(\beta_g d)$ for the H_{313} mode.

$\beta_0 B/k^2 Y_0$, i.e. $\beta_g b_0$. Straight lines of this type are obtained from the results for the H_{313} , H_{412} and H_{411} modes as in Fig. 7, showing the line for the H_{313} mode. The values of k^2 and $B/k^2 Y_0$ can be deduced from these experimental results and are shown in Table 1 together with the theoretical values calculated by the method discussed in Section 7.2.

Table 1

COMPARISON OF EXPERIMENTAL AND THEORETICAL RESULTS FOR EMPTY GROOVES

	Mode					
	H_{313}	H_{411}	H_{412}	H_{114}	H_{214}	H_{121}
k^2 from experiment..	6.6	5.2	5.4	15	8.8	
k^2 from theory ..	6.8	5.5	5.5	17	9.5	193
$B/k^2 Y_0$ from experiment..	0.40	0.19	0.17			-0.09
$B/k^2 Y_0$ from theory ..	0.27	0.16	0.19			

The theoretical results for the H_{313} , H_{411} and H_{412} modes involve the use of a number of approximations, so that the agreement with the experimental values cannot be expected to be particularly good. Nevertheless, the agreement is adequate to confirm that the equivalent circuit in Fig. 2 can represent the behaviour of the cavity-groove combination.

When $1/\delta\lambda$ is plotted against $\beta_g \cot(\beta_g d)$ for the H_{114} and H_{214} modes, the points do not lie on a straight line as predicted by eqn. (9). The results for the H_{114} mode are shown in Fig. 8, where it is seen that instead of a single line there are two separate curves, one for each of the resonances on either side of the critical depth. The reason for this behaviour is that the E_{113} mode resonates at a frequency just above that for the H_{114} mode and its contribution to the susceptance B varies sufficiently rapidly with frequency to invalidate the assumption that B can be considered constant over the range of frequencies for which the H_{114} mode is investigated. Since B is not a constant, it is no longer possible to determine a unique value for it either experimentally or theoretically; the equivalent circuit in Fig. 2 may be modified to meet this situation by including a resonant circuit to represent the E_{113} mode, but, because of the computational labour involved, it has not been considered necessary to seek detailed numerical agreement between theory and experiment. Even when the equivalent circuit is modified, the transformer turns ratio retains its significance and can be calculated as before. The experimental value can be deduced from the straight portions of the curves in Fig. 8, and, as can be seen from Table 1, the agreement between theory and experiment is not much worse than for the other modes. The H_{214} mode shows similar

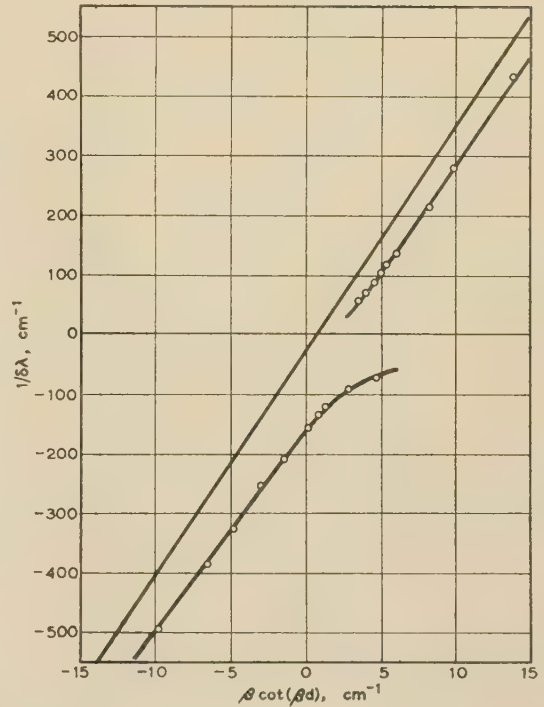


Fig. 8.—Curve of $1/\delta\lambda$ against $\beta_g \cot(\beta_g d)$ for the H_{114} mode.

behaviour because of the close proximity of the resonant frequency of the E_{212} mode.

The frequency shifts for the H_{121} mode are extremely small except near the critical depth of the groove, and it is impossible to draw the line $1/\delta\lambda$ against $\beta_g \cot(\beta_g d)$ with any accuracy. No experimental value for k^2 is therefore available, but $B/k^2 Y_0$ can be deduced from the experimental results by reading the critical depth, d , directly from Fig. 9, in which the transmission

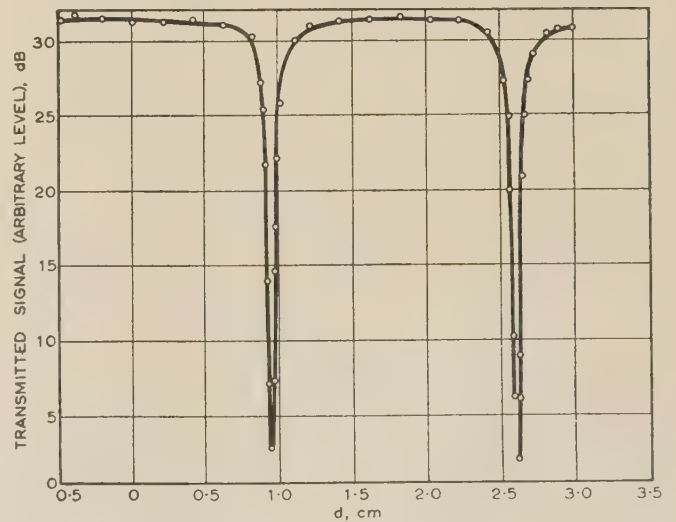


Fig. 9.—Transmission loss for the H_{121} mode caused by an empty groove.

loss is plotted against d , and then using eqn. (19). In the calculation of the theoretical value of $B/k^2 Y_0$ for this mode, the simple expression of eqn. (76) cannot be used, since the H_{11} mode in the main cavity is not evanescent; a modification similar to that

discussed in the previous paragraph is then required to enable the calculation to be completed.

(3.3) Suppression by a Lossy Groove

In Section 2.1 it was shown that the ease with which a mode can be suppressed depends on the coupling between the main cavity and the groove, for which both experimental and theoretical values have been derived in Section 3.3. The most difficult mode to suppress in the cavity in question is the H_{121} , the other modes all being relatively easy to eliminate. Experiments have been carried out using as the lossy material a mixture of iron powder and a cold-setting plastic,⁶ having the values $4.26/-1.1^\circ$ and $1.14/-8.6^\circ$, for the complex relative permittivity and permeability respectively. Measurements were made on the H_{121} mode and the H_{313} mode, the latter being so strongly attenuated near the depth of groove for maximum suppression that observations could be made only for groove depths well removed from the critical value. The observed drop in signal level for the H_{313} mode at these depths agreed quite well with values calculated from the above properties of the lossy material. For a detailed comparison of theoretical and experimental results, it suffices to consider the H_{121} mode.

When the groove is filled with lossy material it changes the resonant frequency and Q -factor of the resonant mode; the former has been adequately dealt with in Section 3.2, so attention can be focused on the latter. The simplest method of determining the change in Q -factor is to measure the change in the insertion loss of the cavity, this also providing a direct check on the actual degree of mode suppression. In the absence of the lossy groove, the loss for transmission through the cavity is given in decibels by

$$L_0 = 20 \log_{10} (1 + Q_{rad}/Q_0) \quad (23)$$

where Q_{rad} is the Q -factor arising from the coupling of the cavity to the external waveguides. Similarly, when the lossy groove is present, the transmission loss in decibels is

$$L_1 = 20 \log_{10} (1 + Q_{rad}/Q) \quad (24)$$

Q being defined by eqn. (11). For cavities which are only loosely coupled to the waveguides, Q_{rad} is much larger than Q_0 , so that the unit factor in the brackets of both eqn. (23) and eqn. (24) can be neglected. Subtraction of L_0 from L_1 and insertion of the expression for Q then gives the additional loss caused by the groove as

$$L_g = L_1 - L_0 = 20 \log_{10} (1 + Q_0/Q_g) \quad (25)$$

The value of Q_g can be calculated from the constants of the material filling the groove by using eqns. (16)–(18).

The experimental results for the H_{121} mode are shown in Fig. 10, in which the loss caused by the groove is plotted against its depth. The general form of the curve is in agreement with the predictions of Section 3.3, a series of values of L_g for maximum suppression occurring at approximately half-wavelength intervals. The wavelength of the H_{10} coaxial mode in the groove, calculated from the material constants given above, is 1.48 cm, when the free-space wavelength has the value 3.24 cm corresponding to the H_{121} resonance. The two depths which give maximum suppression differ by approximately 0.75 cm, in good agreement with the half-wavelength value, 0.74 cm. The first depth for maximum suppression is given by eqn. (19), and calculation using the above value for λ_g and the value from the experiments on the empty groove for $B/k^2 Y_0$ (see Table 1) gives $d = 0.39$ cm as compared with 0.42 cm from Fig. 10.

The loss at the first depth for maximum suppression is considerably greater than that for the second, again in agreement with the discussion in Section 3.3. The theoretical value for

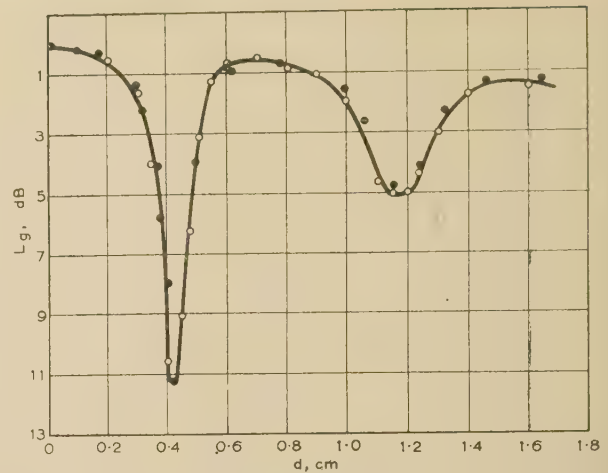


Fig. 10.—Transmission loss for the H_{121} mode caused by a completely filled groove.

Q_g at a point of maximum suppression is $Kg/2$, where K calculated from eqn. (17) is 9.7×10^4 for the H_{121} mode. The points for maximum suppression occur so nearly at the depth equal to odd multiples of $\lambda_g/4$ that g can be taken as $\tanh(\alpha_g d)$ without introducing much error. The attenuation coefficient for the H_{10} mode in the groove, α_g , is found from the loss angles of the material to be 0.36 neper/cm. The only remaining quantity needed for the calculation of L_g is Q_0 , the Q -factor of the empty cylindrical cavity, and this was measured as 18 000. The value of L_g for the first two points of maximum suppression are found from eqn. (25) by inserting the above values, to be 11.5 and 6.0 dB respectively. These are in quite reasonable agreement with the experimental values shown in Fig. 10.

As a further check on the theory, measurements were carried out on the arrangement shown in Fig. 11A, in which the groove

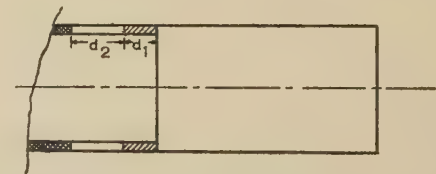


Fig. 11A.—Arrangement of partially filled groove.

was filled to a depth d_1 with the lossy material, and a variable depth d_2 of the groove was left empty between the lossy material and the back plunger. The results are shown in Fig. 11B, where very pronounced peaks of the loss L_g are seen to occur for certain values of the total depth ($d_1 + d_2$). These peaks recur at intervals of half the wavelength measured in the empty groove and show equal values of L_g , since the loss is largely concentrated in the lossy material. Changes in d_2 serve to tune the susceptance term associated with the filled section of the groove, the input admittance of the complete groove then being a pure conductance. The shorter the length of the lossy section, the smaller will be this conductance and the greater the suppression of the mode. This is confirmed by the curves in Fig. 11B. The limiting case occurs when no lossy material is used, the curve for this case being shown in Fig. 9. Normally a double resonance would appear, as discussed earlier, but the H_{121} mode is so loosely coupled to the groove that this does not happen. The empty groove in this case is more effective in suppressing the mode than a lossy groove. Comparison of the curves in Figs. 9, 10 and 11B shows that the

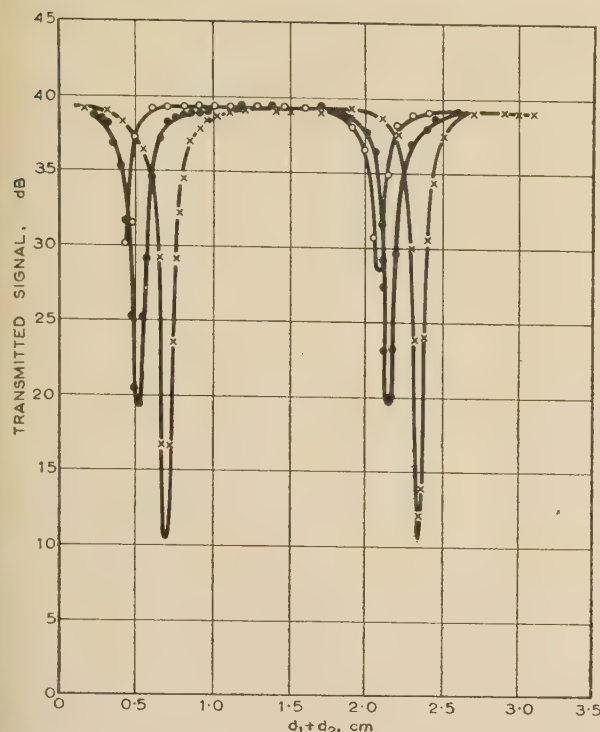


Fig. 11B.—Transmission loss for the H_{121} mode caused by a partially filled groove.

○ $d_1 = 0.4$ cm
● $d_1 = 0.2$ cm
× $d_1 = 0.1$ cm

peaks of L_g become sharper as the loss in the groove is reduced, which means that the design of the suppression system becomes more critical.

(4) GENERAL DISCUSSION

The theoretical and experimental results given in the paper show that the effectiveness of a lossy groove in suppressing unwanted modes in a cylindrical cavity resonator can be predicted reasonably accurately with the help of the equivalent circuit shown in Fig. 2. The essential principle of the method is that the depth of the groove is selected to tune the susceptance arising from the evanescent fields excited in the cavity; the drop in Q-factor under this condition is then more marked the smaller the total loss associated with the groove. If the groove loss becomes too small, however, the original resonance is split and the two resulting peaks may not be adequately suppressed. This suggests that the loss in the groove should always be made sufficiently great to prevent the double resonance. The more nearly this limiting condition is approached, the more critical becomes the groove depth for maximum suppression.

The ease with which the amplitude of an unwanted mode can be reduced depends on the magnitude of the coupling between the cylindrical cavity and the groove. Modes of H_{lmn} type have relatively tight coupling between the cavity and the groove if the indices l and n are large while m is small and can be suppressed quite easily. For such modes no special care need be taken in making the groove depth have its optimum value. Modes with small values of l and n , such as the H_{112} considered in the last paragraph, can be reduced in amplitude only if the depth is made the correct value. It follows, in general, that only one such mode can be effectively dealt with if the groove is to be completely filled with a single material. The use of an arrangement such as that in Fig. 11A gives an extra degree of freedom in the design and

therefore allows two modes to be suppressed. In principle, a succession of layers of materials of different electrical constants suffices for any number of modes, but practical considerations make such an arrangement of doubtful value. An alternative is to use grooves at different positions in the cavity, e.g. in the side walls, and to design each groove to suppress one particular mode. The analysis given here could be modified to provide theoretical results for other positions of the groove.

The complexity of these suggestions for dealing with several modes leads to the conclusion that the most useful application of the lossy-groove technique is as a complement to the suppression achieved by the appropriate choice of coupling holes. The modes which are most easily reduced by the lossy groove are in fact those which cannot readily be eliminated by the coupling system, unless this becomes very complicated.

(5) ACKNOWLEDGMENT

The work described in the paper was carried out in the Electrical Engineering Department of the Imperial College of Science and Technology and the authors are indebted to Dr. Willis Jackson for the facilities there provided.

(6) REFERENCES

- (1) KINZER, J. P., and WILSON, I. G.: 'Some Results on Cylindrical Cavity Resonators', *Bell System Technical Journal*, 1947, **26**, p. 410.
- (2) WILSON, I. G., SCHRAMM, C. W., and KINZER, J. P.: 'High-Q Resonant Cavities for Microwave Testing', *ibid.*, 1946, **25**, p. 408.
- (3) BLEANEY, B., LOUBSER, J. H. N., and PENROSE, R. P.: 'Cavity Resonators for Measurements with Centimetre Electromagnetic Waves', *Proceedings of the Physical Society*, 1947, **59**, p. 185.
- (4) CUNLIFFE, A., and MATHIAS, L. E. S.: 'Some Perturbation Effects in Cavity Resonators', *Proceedings I.E.E.*, Paper No. 1018 R, September, 1950 (**97**, Part III, p. 367).
- (5) SHIMODA, K.: 'Mode Conversion Losses in a TE_{01} -Type Cavity Resonator with Tilted End Plate', *Journal of the Physical Society of Japan*, 1951, **6**, p. 378.
- (6) EL-IBIARY, M. Y.: 'The Microwave Properties of Plastic Loaded with Iron Powder', *Wireless Engineer* (to be published).
- (7) MARCUVITZ, N.: 'Waveguide Handbook', M.I.T. Radiation Laboratory Series No. 10, Sections 2.3 and 2.4 (McGraw-Hill, 1947).
- (8) *Ibid.*, Section 3.5.
- (9) WHITTAKER, E. T., and ROBINSON, G.: 'The Calculus of Observations' (Blackie and Son, Ltd., 1940), p. 136.

(7) APPENDICES

(7.1) Theoretical Treatment

The effect of the groove on the resonances of the main cavity may be calculated by finding what field distributions satisfy the boundary conditions for the structure shown in Fig. 1. Since this structure has axial symmetry, the angular dependence of the fields must be the same in both the main cavity and the groove. It may therefore be assumed that for modes having l cycles in the angular direction the field components E_ϕ and H_ϕ are proportional to $\cos(l\phi)$ and the components E_ρ and H_ρ are proportional to $\sin(l\phi)$. The field in the main cavity must consist of a set of H_{lm} and E_{lm} modes, each mode giving a standing-wave pattern such that E_ϕ and E_ρ vanish on the end wall $z = c$. The

properties of these modes are well known, and the complete fields can be written in the form⁷

$$E_t^I = \sum_{m=1}^{\infty} \left\{ A_m \sin [\beta_{lm}(z - c)] E_m(\rho, \phi) + A'_m \sin [\beta'_{lm}(z - c)] E'_m(\rho, \phi) \right\} \quad (26)$$

$$H_t^I = j n_{\Lambda} \sum_{m=1}^{\infty} \left\{ A_m Y_{lm} \cos [\beta_{lm}(z - c)] E_m(\rho, \phi) + A'_m Y'_{lm} \cos [\beta'_{lm}(z - c)] E'_m(\rho, \phi) \right\} \quad (27)$$

where E_t^I and H_t^I are respectively the electric and magnetic fields transverse to the axis in the main cavity, i.e. region I of Fig. 1, \mathbf{n} is a unit vector in the direction of the cavity axis, A_m is the amplitude of the H_{lm} mode, β_{lm} , Y_{lm} are respectively the phase-change coefficient and wave admittance of the H_{lm} mode, and $E_m(\rho, \phi)$ is a vector describing the variation of the transverse electric field of the H_{lm} mode with ρ and ϕ . The last four quantities when primed refer to the E_{lm} mode. Detailed expressions for the vector functions E_m and E'_m are available in Reference 7; these functions satisfy the orthonormal relations:

$$\left. \begin{aligned} \int_S E_m \cdot E_n dS &= \int_S E'_m \cdot E'_n dS = \delta(m - n) \\ \int_S E_m \cdot E'_n dS &= 0 \end{aligned} \right\} \quad (28)$$

where S is the waveguide cross-section and $\delta(x)$ is the Dirac delta function. The wave admittances, Y_{lm} , satisfy eqn. (7) and the admittances Y'_{lm} , the equation

$$Y'_{lm} = \beta_0 Y_0 / \beta'_{lm} \quad (29)$$

The fields in the groove (region II) may be written as series similar in form to those in eqns. (26) and (27), there being one term for each possible mode; the dominant mode is the H_{11} , and since the groove width, w , is small compared to the radius a , the fields for it can be approximated by⁷

$$E_t^{II} = C \sinh [\gamma_g(z + d)] \cos(l\phi) \mathbf{p}_0 \quad (30)$$

$$H_t^{II} = -Y_g C \cosh [\gamma_g(z + d)] \cos(l\phi) \mathbf{n}_{\Lambda} \mathbf{p}_0 \quad (31)$$

where C is an amplitude constant. To the same order of approximation

$$\gamma_g^2 = (1/a)^2 - \mu_r \epsilon_r \beta_0^2 \quad (32)$$

while Y_g is given by eqn. (8).

On the plane $z = 0$, the transverse electric and magnetic fields must be continuous in the region of the groove, i.e.

$$E_t^I = E_t^{II} \quad \left. \begin{aligned} &\text{for } z = 0 \text{ and } r \leq \rho \leq a \\ &H_t^I = H_t^{II} \end{aligned} \right\} \quad (33)$$

$$(34)$$

Further, on the end wall in the plane $z = 0$ the transverse electric field must vanish, i.e.

$$E_t^I = 0 \text{ for } z = 0 \text{ and } 0 \leq \rho \leq r \quad (35)$$

From the conditions on the electric field in the plane $z = 0$ and the orthonormal relations of eqn. (28), the constants A_m and A'_m can all be expressed in terms of C , with the results that

$$A_m \sin(\beta_{lm}c) = -C \sinh(\gamma_g d) \int_S E_m(\rho, \phi) \cdot \mathbf{p}_0 \cos(l\phi) dS \quad (36)$$

$$A'_m \sin(\beta'_{lm}c) = -C \sinh(\gamma_g d) \int_S E'_m(\rho, \phi) \cdot \mathbf{p}_0 \cos(l\phi) dS \quad (37)$$

where S' is the cross-section of the groove. These equations can be written in the form

$$A_m \sin(\beta_{lm}c) = C f_m \sinh(\gamma_g d) \quad (38)$$

$$A'_m \sin(\beta'_{lm}c) = C f'_m \sinh(\gamma_g d) \quad (39)$$

where f_m, f'_m can be evaluated using the expressions in Reference 7 giving:

$$f_m = \frac{l(\nu_l \pi)^{1/2}}{(\chi_{lm}^2 - l^2)^{1/2} J_l(\chi_{lm})} \int_0^a J_l(\chi_{lm} \rho/a) d\rho \quad (40)$$

$$f'_m = \frac{(\nu_l \pi)^{1/2}}{a J_{l+1}(\chi'_{lm})} \int_r^a J'_l(\chi_{lm} \rho/a) \rho d\rho \quad (41)$$

If the values for the constants A_m, A'_m are inserted in eqn. (27) and the result used in eqn. (34), there results

$$\sum_{m=1}^{\infty} [f_m Y_{lm} E_m(\rho, \phi) \cot(\beta_{lm}c) + f'_m Y'_{lm} E'_m(\rho, \phi) \cot(\beta'_{lm}c)] = -Y_g \cot(\gamma_g d) \cos(l\phi) \mathbf{p}_0 \quad (42)$$

This equation ought to be satisfied for all values of ρ between 0 and a , but in view of the approximation made by neglecting all except the dominant mode in region II, it can only hold exactly at one point. A more complete analysis may be made using Schwinger's variational technique,^{8,9} and it is then found that the best approximation to the final result, namely the condition which gives the resonant frequencies, is obtained by taking the scalar product of both sides of eqn. (42) and the vector $\cos(l\phi) \mathbf{p}_0$ and then integrating over the cross-section of the groove. This gives

$$j \sum_{m=1}^{\infty} [f_m^2 Y_{lm} \cot(\beta_{lm}c) + f'^2_m Y'_{lm} \cot(\beta'_{lm}c)] = Y_g \coth(\gamma_g d) \int_{S'} \cos^2(l\phi) dS \quad (43)$$

The only unknown in eqn. (43) is the frequency, and the possible roots, which form an infinite set, give the resonances of the cavity-groove combination. When d is zero, $\coth(\gamma_g d)$ tends to infinity, and the roots then correspond to the frequencies for which any term $\cot(\beta_{lm}c)$ is infinite, i.e. the H_{lmn} modes of the cylindrical cavity, and for which $\cot(\beta'_{lm}c)$ is infinite, i.e. the E_{lmn} modes. For a lossy-groove filling, the roots become complex and give the resonant frequencies and Q-factors as discussed in Section 2. Exact expressions for these cannot be obtained, but an approximation of the required order results from the procedure already used in Section 2.1. If the change in the resonant properties of, say, the $H_{lm'n}$ mode resulting from the presence of the groove is required, then the appropriate cotangent term in eqn. (43) is replaced by the approximation obtained by using eqns. (4) and (5). The remaining terms in eqn. (43) change relatively slowly with frequency and may be evaluated without appreciable error at the resonant frequency of this mode in the ideal cavity. The shift in the resonant wavelength is therefore given by

$$\frac{-j f'^2_{m'} Y_{lm'} \beta_{lm'} \lambda_0}{c \beta_0^2 \delta \lambda} = \pi a w Y_g \coth(\gamma_g d) - j \sum_{m=1}^{\infty} f_m^2 Y_{lm} \cot(\beta_{lm}c) - j \sum_{m=1}^{\infty} f'^2_m Y'_{lm} \cot(\beta'_{lm}c) \quad (44)$$

where the dash on the first summation indicates the omission of the term for m equal to m' and the integral in eqn. (43) has been evaluated assuming, as before, that w is much less than a .

This equation is similar in form to eqn. (6) and comparison of corresponding terms gives

$$k^2 = \pi a w / (f_m')^2 \quad (45)$$

$$\text{and } f_m'^2 B = - \sum_{m=1}^{\infty} f_m'^2 Y_{lm} \cot(\beta_{lm} c) - \sum_{m=1}^{\infty} f_m'^2 Y'_{lm} \cot(\beta'_{lm} c) \quad (46)$$

The unknown parameters in the equivalent circuit shown in Fig. 2 are thus expressed in terms of the parameters of the modes in the cavity.

(7.2) Evaluation of the Constants of the Equivalent Circuit

A necessary preliminary to the evaluation of k and B is the calculation of the integrals defining f_m and f_m' . Each of these can be expressed exactly as a sum of Bessel functions, but this sum is unsuited for numerical calculations. Accordingly, approximate values are obtained on the assumption that w , the width of the groove, is much less than a . This means that the radius ρ can be taken as having the constant value a in the range from r to a . The expression for f_m' can then be immediately calculated, thus

$$\begin{aligned} f_m' &\simeq \frac{(\nu_l \pi)^{1/2}}{a J_{l+1}(X'_{lm})} \int_r^a J_l(X'_{lm} \rho / a) d\rho \\ &= \frac{a(\nu_l \pi)^{1/2} J_l(X'_{lm} b / a)}{X'_{lm} J'_l(X'_{lm})} \end{aligned} \quad (47)$$

since

$$J_l(X'_{lm}) = 0$$

For large values of X'_{lm} , these Bessel functions can be replaced by their asymptotic expansions; a more convenient approximation in the present case is to observe that the variation of a Bessel function with respect to its argument is similar over a small range to that of a sine or cosine function. In particular, near the value X'_{lm} , the Bessel function $J_l(x)$ is given to a very good approximation by

$$J_l(x) = J'_l(X'_{lm}) \sin(x - X'_{lm}) \quad (48)$$

which ensures both that $J_l(X'_{lm})$ is zero and that $J'_l(x)$ has its correct value at x equal to X'_{lm} . Substitution into eqn. (47) gives

$$f_m' = - \frac{a(\nu_l \pi)^{1/2}}{X'_{lm}} \sin(X'_{lm} w / a) \quad (49)$$

The corresponding approximation for f_m is obtained by starting with the expression similar to eqn. (55) which applies near a root of $J'_l(x) = 0$, namely:

$$J_l(x) = J_l(X_{lm}) \cos(x - X_{lm}) \quad (50)$$

$$\begin{aligned} \text{Then, } f_m &\simeq \frac{l(\nu_l \pi)^{1/2}}{(X_{lm}^2 - l^2)^{1/2}} \int_r^a \cos\left[X_{lm}\left(\frac{\rho}{a} - 1\right)\right] d\rho \\ &= \frac{al(\nu_l \pi)^{1/2} \sin(X_{lm} w / a)}{X_{lm}(X_{lm}^2 - l^2)^{1/2}} \end{aligned} \quad (51)$$

The transformer turns ratio is only required for small values of m and then the Bessel function is virtually constant across the width of the groove. This is equivalent to the approximation $X_{lm} w / a \ll 1$, giving

$$f_m \simeq w l (\nu_l \pi)^{1/2} / (X_{lm}^2 - l^2)^{1/2} \quad (52)$$

$$\text{and } k^2 = a(X_{lm}^2 - l^2) / w \nu_l l^2 \quad (53)$$

$$\text{i.e. } k^2 = (a/w) F_{lm} \quad (54)$$

where $F_{lm} = (X_{lm}^2 - l^2) / \nu_l l^2$ and is tabulated in Table 2.

VOL. 104, PART C.

Table 2

VALUES OF F_{lm}

l	m			
	1	2	3	4
1	1.20	13.7	36.0	68.1
2	0.67	5.11	11.9	21.2
3	0.48	3.08	6.65	
4	0.39	2.19	4.72	

The susceptance B is evaluated by inserting the expressions for f_m and f_m' in eqn. (46). All terms except the first few in each series correspond to evanescent modes for which β_{lm} becomes imaginary. For reasonably large values of m a good approximation to β_{lm} is then found from eqn. (2) to be

$$\beta_{lm} = j X_{lm} / a \quad (55)$$

Further, $\cot(\beta_{lm} c)$ becomes $-j \coth(X_{lm} c / a)$, which can be replaced by $-j$ for large values of m . The m th term of the first series can therefore be approximated as follows:

$$m\text{th term} \simeq \frac{Y_0 \pi a \nu_l l^2 \sin^2(X_{lm} w / a)}{\beta_0 X_{lm} (X_{lm}^2 - l^2)} \quad (56)$$

Since X_{lm}^2 is much larger than l^2 for all except small values of m , the l^2 can be omitted from the denominator, giving

$$m\text{th term} \simeq \pi a \nu_l l^2 Y_0 \sin^2(X_{lm} w / a) / \beta_0 X_{lm}^3 \quad (57)$$

Similarly the m th term of the second series reduces to

$$-\beta_0 Y_0 a^3 \nu_l \pi \sin^2(X'_{lm} w / a) / (X'_{lm})^3 \quad (58)$$

The susceptance B can therefore be evaluated from the expression

$$\begin{aligned} B &= - \frac{\pi a \nu_l Y_0}{(f_m')^2} \left[\sum_{m=1}^M \frac{a l^2 \beta_{lm} \cot(\beta_{lm} c) \sin^2(X_{lm} w / a)}{\beta_0 X_{lm}^2 (X_{lm}^2 - l^2)} \right. \\ &\quad + \frac{l^2}{\beta_0} \sum_{m=M+1}^{\infty} \frac{\sin^2(X_{lm} w / a)}{X_{lm}^3} + \sum_{m=1}^{M'} \frac{\beta_0 a \cot(\beta'_{lm} c) \sin^2(X'_{lm} w / a)}{\beta'_{lm} X_{lm}^2} \\ &\quad \left. - \beta_0 a^2 \sum_{m=M'+1}^{\infty} \frac{\sin^2(X'_{lm} w / a)}{X_{lm}^3} \right] \end{aligned} \quad (59)$$

where M and M' are respectively the numbers of propagating H and E modes. The term for m' is to be excluded from the first sum.

In the two infinite series the Bessel function roots may be approximated by

$$X_{lm} = \left(m + \frac{2l-3}{4}\right) \pi \quad (60)$$

$$X'_{lm} = \left(m + \frac{2l-1}{4}\right) \pi \quad (61)$$

These two series are then closely related to

$$S_l(w/a) = \frac{1}{\pi^3} \sum_{m=1}^{\infty} \frac{\sin^2\left[\left(m + \frac{2l-3}{4}\right) \pi w / a\right]}{\left(m + \frac{2l-3}{4}\right)^3} \quad (62)$$

and
$$S'_l(w/a) = \frac{1}{\pi^3} \sum_{m=1}^{\infty} \frac{\sin^2 \left[\left(m + \frac{2l-1}{4} \right) \pi w/a \right]}{\left(m + \frac{2l-1}{4} \right)^3} \quad (63)$$

It is easily verified that

$$S_{l+2}(w/a) = S_l(w/a) - \frac{\sin^2 [(2l+1)\pi w/4a]}{[(2l+1)/4]^3} \quad (64)$$

and
$$S'_l(w/a) = S'_{l+1}(w/a) \quad (65)$$

so that only $S_1(w/a)$ and $S_2(w/a)$ need be evaluated. To the order of approximation implied by eqns. (60) and (61),

$$\sum_{m=M+1}^{\infty} \frac{\sin^2 (X'_{lm} w/a)}{\chi_{lm}^3} = S_{l+2M}(w/a) \quad (66)$$

$$\sum_{m=M'+1}^{\infty} \frac{\sin^2 (X'_{lm} w/a)}{\chi_{lm}^3} - S'_{l+2M'}(w/a) = S_{l+2M'+1}(w/a) \quad (67)$$

For small values of w/a , the series $S_l(w/a)$ can be evaluated to a reasonable degree of accuracy by the Euler-Maclaurin formula⁹ with the result

$$S_l(w/a) \simeq \frac{1}{2\pi} \left(\frac{w}{a} \right)^2 \left\{ \log_e [a/(2l+1)w] + 1.97 + \frac{4}{2l+1} + \frac{1.5}{(2l+1)^2} \right\} \quad (68)$$

The simplest application of eqn. (66) is to an H_{l1n} mode with no propagating E_{lm} modes, when

$$\frac{B}{N^2 Y_0} = - \frac{2}{(\beta_0 w)} [l^2 S_{l+2}(w/a) - (\beta_0 a)^2 S_{l+1}(w/a)] \quad (69)$$

HALL EFFECT AND ITS COUNTERPART, RADIATION PRESSURE, IN MICROWAVE POWER MEASUREMENT

By Professor H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), Member.

(The paper was first received 4th April, and in revised form 23rd May, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

Hall effect and radiation pressure are shown to be different manifestations of the same basic phenomenon, and their relationship is deduced for a plane wave transmitted through a slab of material of given electrical characteristics. In particular, the contributions to these effects of both the conduction and displacement components of current in the material are examined, and it is shown that the latter can be a significant supplementary factor at ultra-high frequencies.

Both Hall effect and radiation pressure have been used for microwave power measurement, and some aspects of these applications are discussed in the light of the present analysis. To permit a better assessment of the Hall-effect-wattmeter problem, numerical calculations have been made at 4000 Mc/s for *n*-type germanium and silicon showing that a high-resistivity material is necessary to achieve a satisfactory performance with a transmission type of instrument. Similar computations for polythene predict that a measurable Hall effect arising from displacement current in this material should exist, and it is suggested that observations based on that effect might offer a new approach in further exploration of the properties of dielectrics.

LIST OF PRINCIPAL SYMBOLS

E_s, H_s = Resultant electric and magnetic field components at $x = s$.

E_x, E_y = Components of electric field in the directions x and y .

E_{xc} or E_c, E_{xd} or E_d = Electric field of the Hall effect arising from conduction and displacement currents respectively.

E_0, H_0 = Incident electric and magnetic field components at $x = 0$ just inside the slab.

f_{xc}, f_{xd} = Mechanical force accompanying radiation pressure in the direction x and arising from conduction and displacement currents respectively.

H_s^* = Complex conjugate of H_s .

H_z, B_z = Magnetic field strength and magnetic flux density respectively in the direction z .

I_y, J_y = Total current and current density respectively in the direction y .

J_{yc}, J_{yd} = Conduction and displacement current densities respectively in the direction y .

N_c = Number of mobile carriers per unit volume.

N_d = Number of electric dipoles per unit volume.

$P_{(av)} = P_{c(av)} + P_{d(av)}$ = Time average of the resultant radiation pressure per unit area.

$P_c, P_{c(av)}$ = Radiation pressure per unit area arising from conduction current, and time average of corresponding quantity.

$P_d, P_{d(av)}$ = Radiation pressure per unit area arising from displacement current, and time average of corresponding quantity.

P_s = Power density at $x = s$.

P_1 = Power density at $x = 0$.

q_c = Charge on each mobile carrier.

q_d = Charge per pole of each electric dipole.

\mathcal{R}_c = 'Hall coefficient' associated with conduction current.

\mathcal{R}_d = 'Hall coefficient' associated with displacement current.

s = Thickness in the direction x of slab of the medium of interest.

$V_{(av)} = V_{c(av)} + V_{d(av)}$ = Time average of the resultant Hall e.m.f.

$V_c, V_{c(av)}$ = Hall e.m.f. arising from conduction current, and time average of corresponding quantity.

$V_d, V_{d(av)}$ = Hall e.m.f. arising from displacement current, and time average of corresponding quantity.

x, y, z = Cartesian co-ordinates.

$Z = |Z| \angle \delta$ = Intrinsic impedance of the slab.

$Z_s = |Z_s| \angle \phi$ = Load impedance.

α = Attenuation coefficient.

β = Phase-change coefficient.

$\gamma = |\gamma| \angle \psi$ = Propagation coefficient.

ϵ_r = Relative permittivity.

ϵ_0, μ_0 = Permittivity and permeability, respectively, of free space.

$\rho = |\rho| \angle \theta$ = Voltage reflection coefficients at $x = s$.

σ, ϵ, μ = Conductivity, permittivity and permeability, respectively, of the medium of interest.

(1) INTRODUCTION

As long ago as 1880, Hall¹ found that the streamlines of an electric current in a metallic conductor were distorted when the conductor was placed in an external magnetic field. This distortion arises from an electric field set up at right angles to the current and to the magnetic field. The effect is well known in conductors and semi-conductors, both for steady and low-frequency alternating currents.² In these circumstances, calculations have naturally been based upon the interaction of the external magnetic field with a current depending upon the electrical conductivity of the medium. So far as the author is aware, no account has been taken of displacement current, and, indeed, it is only at very high frequencies in non-metallic media that contributions from such currents can be significant. Donovan³ has recently developed a theory, neglecting displacement currents, of the high-frequency Hall effect in metals arising from a constant independently applied magnetic field. The problem treated in the present paper is quite different, because

Correspondence on Monographs is invited for consideration with a view to publication.
Professor Barlow is Pender Professor of Electrical Engineering at University College, London.

it concerns the effect produced solely by the field components of a high-frequency electromagnetic wave.

It is the purpose here to examine theoretically on a purely classical basis the possibilities in the development of a Hall effect at high frequencies in media for which displacement currents might be expected to make a finite contribution. In this analysis it will be apparent that radiation pressure is just another manifestation of the same basic phenomenon and that the two effects are therefore intimately related to one another. Both effects have been used for microwave power measurement, and some of the factors that determine their relative merits for this purpose are discussed.

(2) BASIC PRINCIPLES

Using Cartesian co-ordinates (Fig. 1), suppose a current I_y flows at right angles to a magnetic field of density B_z ; then the mechanical force on the current is given by

$$f_x = B_z I_y \quad (1)$$

per unit length of the current path.

This force acts on the electric charges, which in moving under the influence of the electric field applied along the y -axis, con-

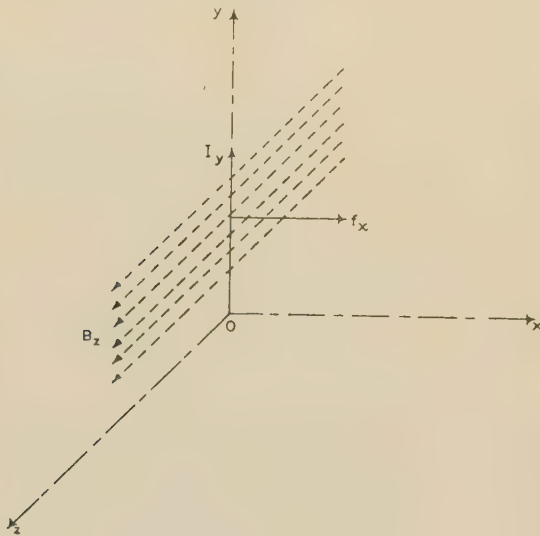


Fig. 1.—Co-ordinate system.

stitute the current. The whole of the force f_x must be transferred to the material medium associated with these charges, and in that process the current streamlines will be pushed sideways in the direction x , setting up a secondary electric field, E_x . This mechanism is well established and it enables us on the one hand to calculate the force on the conductors of an electric motor and on the other to evaluate the radiation pressure when an electromagnetic wave is incident on a surface. The secondary electric field E_x , which is inherent in the mechanism, represents the so-called 'Hall effect' and, as has been pointed out,⁴ radiation pressure is really dependent on the existence of a Hall effect. J. J. Thomson calculated the radiation pressure on a metallic conductor using the relation (1), the current being induced in the surface by the electromagnetic wave. The result has been verified by Nichols and Hull⁵ for light waves and by Cullen⁶ for microwaves.

For semi-conductors and imperfect dielectrics in which displacement currents as well as conduction currents exist, it is to be expected that a mechanical force, f_x , and a corresponding secondary electric field, E_x , will be produced to operate upon the

movable charges representing both components of the current in the material medium. Some of these charges will be mobile, passing from one molecule to another under the action of the primary applied electric field, while others will be associated with electric dipoles and will suffer only local displacement.

To simplify the problem we shall confine attention to media in which the movable charges are overwhelmingly electrons; thus n -type semi-conductors would be included and so would good high-frequency dielectrics like polythene, in which it is known that the positive charges associated with the atomic nuclei have so large an inertia that they are unable to respond to rapidly alternating forces and take no appreciable part in the support of a displacement current. Substances which include polar molecules would not comply with the requirements postulated, but an extension of the treatment could be made to meet their case.

(3) ELEMENTARY THEORY

We shall assume a plane wave propagated in air or a vacuum and incident in the direction x on the medium of interest, whose surface lies in the yz -plane at $x = 0$ (see Fig. 2). The field components of the wave are E_y and H_z , to which the following relationships apply:

$$-\frac{\partial H_z}{\partial x} = \sigma E_y + \epsilon \frac{\partial E_y}{\partial t} = J_y \quad (2)$$

$$B_z = \mu H_z \quad (3)$$

The total current density, $J_y = J_{yc} + J_{yd}$, includes both the conductivity and displacement components.

In considering the mechanical forces and associated electric fields arising from the interaction of these two component currents with the magnetic field of the wave, it must be recognized that the mechanism is concerned in the one case with mobile charges and in the other with bound charges. It is therefore appropriate to deal with each current separately and subsequently to combine their effects. Thus, with reference to Fig. 2 for an elemental volume dx, dy, dz of the medium, the mechanical force df_{xc} on the conductivity component of current ($J_{yc} dx dz$) in the element over the length dy is

$$df_{xc} = B_z (J_{yc} dx dz) dy \quad (4)$$

where

$$J_{yc} = \sigma E_y \quad (5)$$

The corresponding expression for the displacement component of current ($J'_{yd} dx dz$) must take into account the fact that only the material medium can contribute in this respect, otherwise we should have a force of the same kind in a vacuum.

Therefore

$$df_{xd} = B_z (J'_{yd} dx dz) dy \quad (6)$$

where

$$J'_{yd} = (\epsilon - \epsilon_0) \frac{\partial E_y}{\partial t} \quad (7)$$

Now, as already pointed out, the force f_{xc} acts on the mobile charges in the medium and is associated with an electric field E_{xc} , so that we can write

$$df_{xc} = m q_c N_c E_{xc} dx dy dz \quad (8)$$

where m is a factor to take into account the fact that the electric force on the individual charges differs from the applied field. Similarly, since f_{xd} acts on electric dipoles and is associated with an electric field E_{xd} , we have

$$df_{xd} = n q_d N_d E_{xd} dx dy dz \quad (9)$$

where n is a factor corresponding to m for the mobile charges. Dropping the co-ordinate subscripts and combining eqns. (3), (4), (5) and (8) gives

$$m q_c N_c E_c = \mu H \sigma E \quad (10)$$

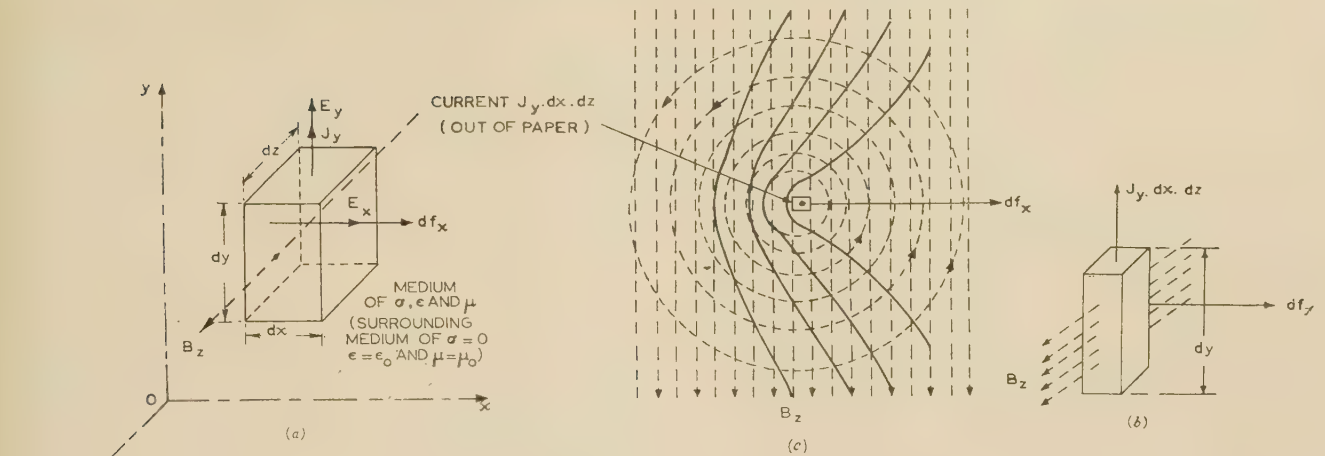


Fig. 2.—Hall-effect electric field and radiation pressure for an elementary volume of the medium.

(a) Element of medium.
(b) Element of medium.
(c) Resultant magnetic field distribution.

and, again from eqns. (3), (6), (7) and (9), we get

$$nq_d N_d E_d = \mu H (\epsilon - \epsilon_0) \frac{\partial E}{\partial t} \quad (11)$$

Thus the electric field of the Hall effect arising from the conductivity of the medium is, from eqn. (10),

$$E_c = \left(\frac{1}{mq_c N_c} \right) \mu H \sigma E = \mathcal{R}_c \mu \sigma E H \quad (12)$$

and the corresponding electric field arising from displacement in the material medium is, from eqn. (11),

$$E_d = \left(\frac{1}{nq_d N_d} \right) \mu H (\epsilon - \epsilon_0) \frac{\partial E}{\partial t} = \mathcal{R}_d \mu (\epsilon - \epsilon_0) H \frac{\partial E}{\partial t} \quad (13)$$

where

$$\mathcal{R}_c = \left(\frac{1}{mq_c N_c} \right) \quad (14)$$

and

$$\mathcal{R}_d = \left(\frac{1}{nq_d N_d} \right) \quad (15)$$

For semi-conductors $m = 8/3\pi$ and for n -type germanium, in which electrons far outnumber holes as mobile carriers, we therefore have

$$\mathcal{R}_c = \frac{3\pi}{8q_c N_c}$$

in accordance with established theory.

As a rough approximation we get the value of n from the Clausius-Mossotti relation as

$$n \simeq \left(\frac{\epsilon_r + 2}{3} \right)$$

For the radiation pressure, representing the force per unit surface area, we use eqns. (8) and (9), giving

$$dp_c = mq_c N_c E_c dx = \frac{1}{\mathcal{R}_c} E_c dx \quad (16)$$

and

$$dp_d = nq_d N_d E_d dx = \frac{1}{\mathcal{R}_d} E_d dx \quad (17)$$

(4) CALCULATION OF THE HIGH-FREQUENCY HALL E.M.F. AND THE CORRESPONDING RADIATION PRESSURE FOR A MEDIUM OF FINITE THICKNESS

For a slab of the medium of thickness s in the x -direction (Fig. 3) the Hall e.m.f.'s arising from conduction and displacement currents respectively are

$$V_c = \int_0^s E_c dx = \mathcal{R}_c \mu \sigma \int_0^s E H dx \quad (18)$$

and

$$V_d = \int_0^s E_d dx = \mathcal{R}_d \mu \epsilon_0 (\epsilon_r - 1) \int_0^s H \frac{\partial E}{\partial t} dx \quad (19)$$

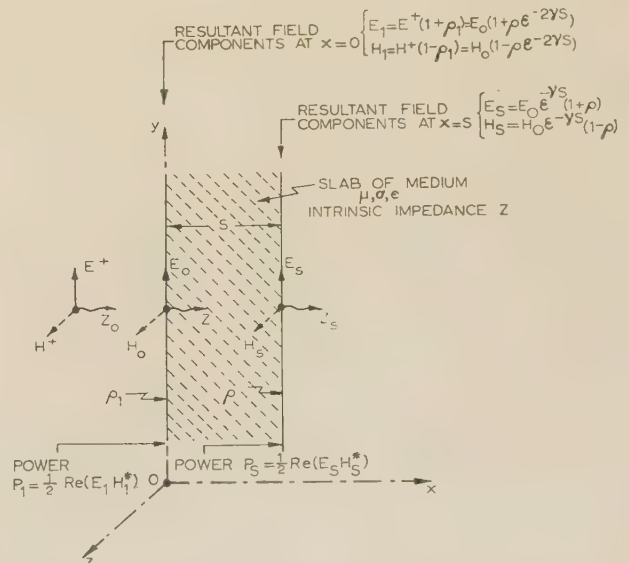


Fig. 3.—Field components, impedances, reflection coefficients and power densities in the path of the wave.

The corresponding radiation pressures per unit surface area are, from eqns. (16) and (17),

$$p_c = V_c / \mathcal{R}_c \quad . \quad . \quad . \quad (20)$$

and

$$p_d = V_d / \mathcal{R}_d \quad . \quad . \quad . \quad (21)$$

It is clear therefore that the Hall coefficients \mathcal{R}_c and \mathcal{R}_d are given by the ratio between the relevant Hall e.m.f. and the accompanying radiation pressure.

Assuming that the field components within the slab vary with distance x according to the function $e^{\pm \gamma x}$

$$\text{where } \gamma = |\gamma| e^{j\psi} = \alpha + j\beta = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} \quad . \quad (22)$$

we can write for the resultant field components at any point between $x = 0$ and $x = s$

$$E = E_0 [e^{-\gamma x} + \rho e^{-\gamma(2s-x)}] \quad . \quad . \quad (23)$$

and

$$H = H_0 [e^{-\gamma x} - \rho e^{-\gamma(2s-x)}] \quad . \quad . \quad (24)$$

so that the intrinsic impedance of the medium under consideration is given by

$$E_0 / H_0 = Z = |Z| e^{j\delta} = \frac{j\omega\mu}{\gamma} \quad . \quad . \quad (25)$$

and the voltage reflection coefficient at the surface where $x = s$ has the value ρ .

The components of the resultant field at $x = s$ are

$$\begin{aligned} E_s &= E_0 e^{-\gamma s} (1 + \rho) \\ H_s &= H_0 e^{-\gamma s} (1 - \rho) \end{aligned} \quad . \quad . \quad (26)$$

giving an impedance beyond that point, representing the load, as

$$Z_s = |Z_s| e^{j\phi} = E_s / H_s \quad . \quad . \quad (27)$$

Thus

$$\rho = |\rho| e^{j\theta} = \frac{Z_s - Z}{Z_s + Z} \quad . \quad . \quad (28)$$

Assuming that the field varies sinusoidally in time, so that $E_0 = |E_0| e^{j\omega t}$ and from eqn. (25), $H_0 = |H_0| e^{j(\omega t - \delta)}$, we can use eqns. (23) and (24) to carry out the integrations of eqns. (18) and (19) to yield

$$V_c = \frac{\mathcal{R}_c \mu \sigma |E_0| |H_0| e^{j(2\omega t - \delta)}}{2\gamma} (1 - e^{-2\gamma s}) (1 - \rho^2 e^{-2\gamma s}) \quad (29)$$

and

$$V_d = \frac{\mathcal{R}_d \mu \omega \epsilon_0 (\epsilon_r - 1) |E_0| |H_0| e^{j(2\omega t - \delta)}}{2\gamma} (1 - e^{-2\gamma s}) (1 - \rho^2 e^{-2\gamma s}) \quad . \quad . \quad (30)$$

and since

$$\frac{|E_0| |H_0|}{2} = \frac{|E_s|^2}{2|Z_s| e^{-2\alpha s} \sqrt{(1 + |\rho|^4 - 2|\rho|^2 \cos 2\theta)}} = \frac{|E_s|^2 \cos \phi}{2|Z_s| e^{-2\alpha s} [(1 - |\rho|^2) \cos \delta - 2|\rho| \sin \delta \sin \theta]} \quad . \quad . \quad (38)$$

with

$$\alpha = |\gamma| \sin \delta \quad \text{and} \quad \beta = |\gamma| \cos \delta$$

we can rewrite eqns. (33) and (34) as

$$V_{c(av)} = \left[\frac{\mathcal{R}_c \mu \sigma}{2} \right] \left[\frac{\frac{1}{\alpha} (e^{2\alpha s} - 1) (1 - |\rho|^2 e^{-2\alpha s}) \cos \delta - \frac{1}{\beta} (2|\rho| \sin \delta) (\sin \theta \sin 2\beta s + \cos \theta \cos 2\beta s - \cos \theta)}{(1 - |\rho|^2) \cos \delta - 2|\rho| \sin \delta \sin \theta} \right] P_s \quad . \quad (39)$$

$$\text{and } V_{d(av)} = \left[\frac{\mathcal{R}_d \mu \omega \epsilon_0 (\epsilon_r - 1)}{2} \right] \left[\frac{\frac{1}{\beta} (1 - e^{2\alpha s}) (1 - |\rho|^2 e^{-2\alpha s}) \cos \delta - \frac{1}{\alpha} (2|\rho| \sin \delta) (\sin \theta \sin 2\beta s + \cos \theta \cos 2\beta s - \cos \theta)}{(1 - |\rho|^2) \cos \delta - 2|\rho| \sin \delta \sin \theta} \right] P_s \quad . \quad (40)$$

In employing the Hall effect or the accompanying radiation pressure to measure power we require the average value in time evaluated for each element of length dx and then integrated over the total thickness s . This procedure is required in recognition of the fact that there is a progressive change in the phase of E with respect to H as x increases.

$$\text{Hence } V_{c(av)} = \mathcal{R}_c \mu \sigma \int_0^s \frac{1}{2} \text{Re}(H^* E) dx \quad . \quad . \quad (31)$$

$$\text{and } V_{d(av)} = \mathcal{R}_d \mu \epsilon_0 (\epsilon_r - 1) \int_0^s \frac{1}{2} \text{Re} \left(H^* \frac{\partial E}{\partial t} \right) dx \quad (32)$$

where Re denotes the real part of the quantity in parentheses, and the calculation is based on maximum values in time of the field components.

Applying eqns. (23) and (24) to the solution of eqns. (31) and (32) gives

$$\begin{aligned} V_{c(av)} &= \frac{\mathcal{R}_c \mu \sigma |E_0| |H_0|}{4|\gamma|} \left\{ \frac{\beta}{\alpha} (1 - e^{-2\alpha s}) (1 - |\rho|^2 e^{-2\alpha s}) \right. \\ &\quad \left. + \frac{2\alpha |\rho| e^{-2\alpha s}}{\beta} [\cos \theta - \cos (2\beta s - \theta)] \right\} \quad . \quad (33) \end{aligned}$$

and

$$\begin{aligned} V_{d(av)} &= \frac{\mathcal{R}_d \mu \omega \epsilon_0 (\epsilon_r - 1) |E_0| |H_0|}{4|\gamma|} \left\{ (e^{-2\alpha s} - 1) (1 - |\rho|^2 e^{-2\alpha s}) \right. \\ &\quad \left. + 2|\rho| e^{-2\alpha s} [\cos \theta - \cos (2\beta s - \theta)] \right\} \quad . \quad (34) \end{aligned}$$

The total Hall e.m.f. therefore has an average value in time of

$$V_{(av)} = V_{c(av)} + V_{d(av)} \quad . \quad . \quad (35)$$

In the same way the time average of the radiation pressure arising jointly from conduction and displacement currents is

$$p_{(av)} = p_{c(av)} + p_{d(av)} = \frac{V_{c(av)}}{\mathcal{R}_c} + \frac{V_{d(av)}}{\mathcal{R}_d} \quad . \quad (36)$$

We are particularly interested in interpreting eqns. (35) and (36) in relation to the impedance, Z_s , of the load and the field at that point.

The power density at $x = s$ and dissipated by the load is

$$P_s = \frac{1}{2} \text{Re}(E_s H_s^*) = \frac{|E_s|^2 \cos \phi}{2|Z_s|} \quad . \quad . \quad (37)$$

When $\alpha s \ll 1$ and $\beta s \ll 1$ we can simplify expression (39) by putting

$\sin 2\beta s \simeq 2\beta s$, $\cos 2\beta s \simeq 1$, $(\epsilon^{2\alpha s} - 1) \simeq 2\alpha s$ and $\epsilon^{-2\alpha s} \simeq 1$ so that

$$V_{c(av)} \simeq (\mathcal{R}_c \mu \sigma s)(P_s) \quad (41)$$

giving the kind of direct relationship between Hall e.m.f. and power which we require to establish as closely as possible. It is also of interest to determine the input power density, P_1 , at $x = 0$, and this is found to be

$$P_1 = \frac{\epsilon^{2\alpha s}(1 - |\rho|^2 \epsilon^{-4\alpha s}) \cos \delta - (2|\rho| \sin \delta)(\sin \theta \cos 2\beta s - \cos \theta \sin 2\beta s)}{(1 - |\rho|^2) \cos \delta - 2|\rho| \sin \delta \sin \theta} P_s \quad (42)$$

The power lost in passing through unit area of the medium concerned is $(P_1 - P_s)$.

(5) NUMERICAL VALUES FOR *n*-TYPE GERMANIUM, *n*-TYPE SILICON AND POLYTHENE

In order to examine the behaviour of different media expected to exhibit Hall effect and radiation pressure under the conditions postulated, a number of calculations have been made for a frequency of 4000 Mc/s using the expressions deduced.

(5.1) *n*-Type Germanium

The germanium is considered to have the following constants:

Resistivity = 26 ohm-cm ($\sigma = 3.85$ mhos/m).

$\epsilon_r = 16$

$\mu = \mu_0 = 4\pi \times 10^{-7}$ henry/m

From these figures we find

$\alpha = 163$ nepers/m

$\beta = 372$ rad/m

$|\gamma| = 406$ per metre

$|Z| = 77.8$ ohms

$\delta = 23.7^\circ$

Taking the density of germanium as 5.3 and using Avogadro's number $N_0 = 6 \times 10^{23}$ we estimate that this semi-conductor contains about 4.5×10^{25} molecules per cubic metre; and since the structure suggests four electric dipoles per molecule, each with the electronic charge 1.6×10^{-19} coulomb, we have $\mathcal{R}_d = 1/nq_d N_d$ with $n \simeq (\epsilon_r + 2)/3 = 6$ giving $\mathcal{R}_d = 5.8 \times 10^{-9}$ m³/coulomb.

From independent measurements we know that $\mathcal{R}_c = 7.4 \times 10^{-3}$ m³/coulomb corresponding to $N_c = 10^{21}$ electrons per cubic metre. Applying these numerical values to calculate the quantity in the first square bracket of each of the eqns. (39) and (40) we find

$$\left[\frac{\mathcal{R}_c \mu \sigma}{2} \right] = 1.79 \times 10^{-8}$$

$$\text{and} \quad \left[\frac{\mathcal{R}_d \mu \omega \epsilon_0 (\epsilon_r - 1)}{2} \right] = 1.2 \times 10^{-14}$$

It is therefore clear that, since the remaining parts of eqns. (39) and (40) cannot be very different in magnitude, the contribution of $V_{d(av)}$ to $V_{(av)}$ is relatively insignificant in this case and can be neglected. It must, however, be emphasized that this result depends on the comparatively small value of \mathcal{R}_d which the above computation yields. If \mathcal{R}_d proved, in fact, to be somewhat larger than the value calculated, the displacement contribution to $V_{(av)}$ would no longer be negligible.

Taking the figures given as correct we can determine $V_{(av)} \simeq V_{c(av)}$ from eqns. (37) and (39) for a given thickness s of the slab, a given load impedance $Z_s = |Z_s| e^{j\phi}$ and a given electric field $|E_s|$. Eqn. (28) relates the reflection coefficient ρ to the load impedance Z_s . The curves in Fig. 4 show values of

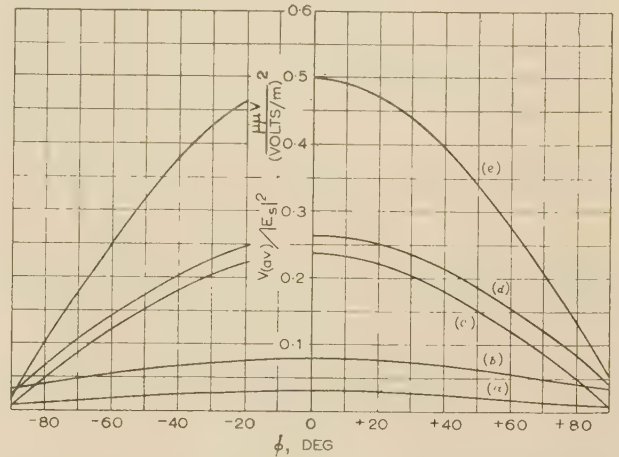


Fig. 4.—Hall e.m.f., $V_{(av)}$, for a given electric field strength E_s and different load impedances $Z_s = |Z_s| e^{j\phi}$ for 26 ohm-cm *n*-type germanium at 4000 Mc/s.

- (a) $|Z_s| = 377\Omega$, $S = 0.05$ cm.
- (b) $|Z_s| = 377\Omega$, $S = 0.1$ cm.
- (c) $|Z_s| = 38.9\Omega$, $S = 0.05$ cm.
- (d) $|Z_s| = 77.8\Omega$, $S = 0.1$ cm.
- (e) $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.

$V_{(av)}/|E_s|^2$ for different load impedances, and it will be observed that there is a substantial departure from the ideal form represented by $\cos \phi$. Not only are the curves asymmetrical about the unity-power-factor axis, but they are also lifted up so that at a phase angle of 90° there is a positive intercept which is a measure of the loss in the slab under these conditions. For a power of 10 watts/cm² consumed by a load $Z_s = 38.9/0^\circ$ ohms with $s = 0.1$ cm, Fig. 4 gives $V_{(av)}/|E_s|^2 = 0.5 \times 10^{-12}$ and hence $V_{(av)} = 3.89 \mu\text{V}$.

Fig. 5 emphasizes the inaccuracy that would arise if germanium were applied in this way to form a transmission type of wattmeter with the power passing through the slab. For satisfactory operation $V_{(av)}/P_s$ should remain constant for all values of Z_s ,

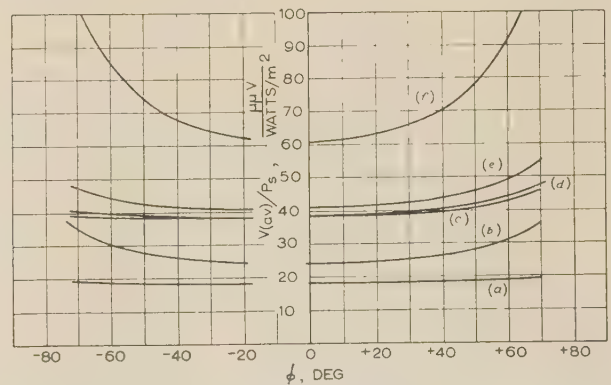


Fig. 5.—Hall e.m.f., $V_{(av)}$, for a given power P_s dissipated in loads of different impedances, $Z_s = |Z_s| e^{j\phi}$, for 26 ohm-cm *n*-type germanium at 4000 Mc/s.

- (a) $|Z_s| = 38.9\Omega$, $S = 0.05$ cm.
- (b) $|Z_s| = 377\Omega$, $S = 0.05$ cm.
- (c) $|Z_s| = 5\Omega$, $S = 0.1$ cm.
- (d) $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.
- (e) $|Z_s| = 77.8\Omega$, $S = 0.1$ cm.
- (f) $|Z_s| = 377\Omega$, $S = 0.1$ cm.

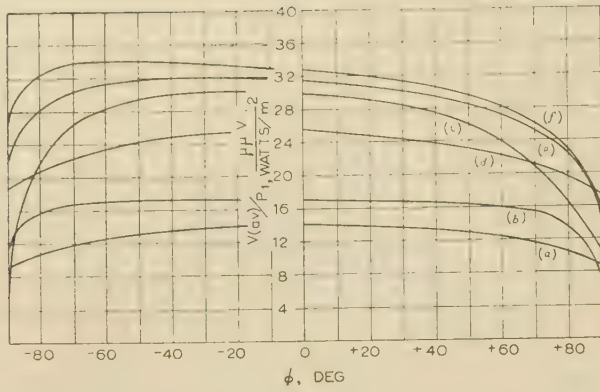


Fig. 6.—Hall e.m.f., $V_{(av)}$, for a given power input, P_1 , and loads of different impedances, $Z_s = |Z_s|/\phi$, for 26 ohm-cm n -type germanium at 4000 Mc/s.

- (a) $|Z_s| = 377\Omega$, $S = 0.05$ cm.
- (b) $|Z_s| = 38.9\Omega$, $S = 0.05$ cm.
- (c) $|Z_s| = 5\Omega$, $S = 0.1$ cm.
- (d) $|Z_s| = 377\Omega$, $S = 0.1$ cm.
- (e) $|Z_s| = 77.8\Omega$, $S = 0.1$ cm.
- (f) $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.

in both magnitude and phase. The error depends very much on the losses in the slab, which are accentuated when $|Z_s|$ and s are increased. Fig. 6 shows the same family of curves as Fig. 5, but calculated in relation to the power input. It is of interest to plot the power on both sides of the slab for a given electric field strength, and by taking the difference to determine the power lost in the slab. This has been done for various load impedances, and the result is given in Fig. 7.

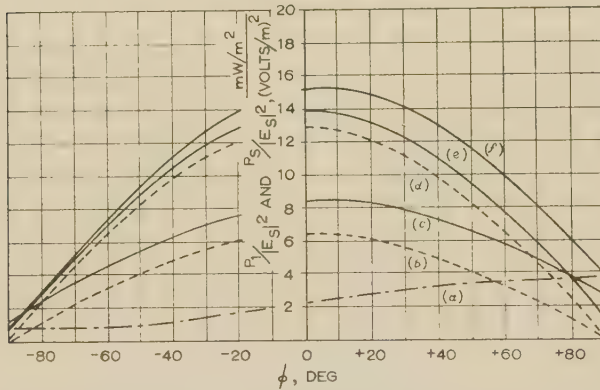


Fig. 7.—Power input, P_1 , and power dissipated, P_s , in the load, for a given electric field, E_s , and different load impedances, $Z_s = |Z_s|/\phi$, for 26 ohm-cm n -type germanium at 4000 Mc/s.

- (a) $(P_1 - P_s)/E_s^2$ for $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.
- (b) P_s/E_s^2 for $|Z_s| = 77.8\Omega$, $S = 0.1$ cm.
- (c) P_1/E_s^2 for $|Z_s| = 77.8\Omega$, $S = 0.1$ cm.
- (d) P_s/E_s^2 for $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.
- (e) P_1/E_s^2 for $|Z_s| = 38.9\Omega$, $S = 0.05$ cm.
- (f) P_1/E_s^2 for $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.

The two components of radiation pressure arising from the conduction and displacement currents are approximately of the same order of magnitude, their values being plotted in Fig. 8. $P_{c(av)}$ must always be positive, since with a finite loss in the slab, power will be absorbed whatever the nature of the load. On the other hand, $P_{d(av)}$, which is dependent upon displacement current, may be either positive or negative in accordance with the condition that this pressure will be exerted in such a way as to tend to reduce the energy stored in the field.

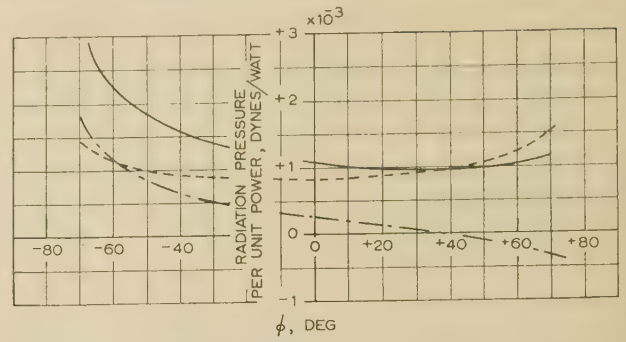


Fig. 8.—Radiation pressure arising from conduction, $P_{c(av)}$, displacement, $P_{d(av)}$, and from total, $p(av)$, currents, for a given power, P_s , dissipated in a load of various phase angles, ϕ , for 26 ohm-cm n -type germanium at 4000 Mc/s.

$$|Z_s| = 377\Omega, S = 0.1 \text{ cm.}$$

$$\begin{aligned} \text{---} & P_{d(av)}/P_s \\ \text{- - -} & P_{c(av)}/P_s \\ \text{---} & p(av)/P \end{aligned}$$

(5.2) n -Type Silicon

The intrinsic resistivity of silicon is 63 600 ohm-cm, and as a semi-conductor it has been produced with resistivities as high as 1000 ohm-cm or more. For power measurement by means of the Hall effect a high-resistivity medium has obvious advantages, and calculations have therefore been made for silicon as follows:

$$\text{Resistivity} = 500 \text{ ohm-cm } (\sigma = 0.2 \text{ mho/m})$$

$$\epsilon_r = 12$$

$$\mu = \mu_0 = 4\pi \times 10^{-7} \text{ henry/m}$$

At 4000 Mc/s these figures yield

$$\alpha = 10.9 \text{ nepers/m}$$

$$\beta = 290 \text{ rad/m}$$

$$|\gamma| = 290 \text{ per metre}$$

$$|Z| = 108.8 \text{ ohms}$$

$$\delta = 2.15^\circ$$

with an estimated $\mathcal{R}_c = 6.2 \times 10^{-2} \text{ m}^3/\text{coulomb}$ and an \mathcal{R}_d of the same order as the value obtained for germanium. Employing these figures to evaluate $V_{(av)}/E_s^2$ and $V_{(av)}/P_s$ for silicon we get Figs. 9 and 10 respectively. It is clear that with this material used in a Hall-effect wattmeter of the transmission type there is a great improvement in performance and the ideal conditions are closely approached except for loads that are almost purely reactive.

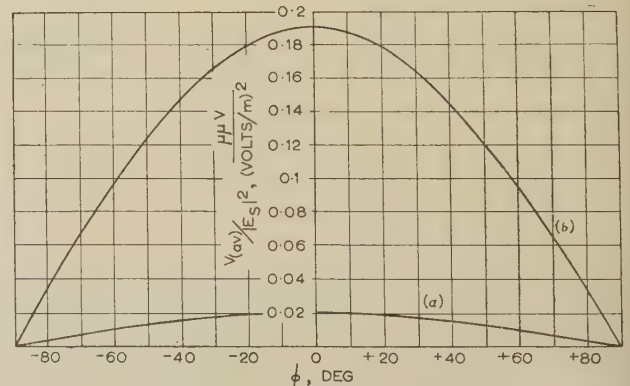


Fig. 9.—Hall e.m.f., $V_{(av)}$, for a given electric field, E_s , and different load impedances, $Z_s = |Z_s|/\phi$, for 500 ohm-cm n -type silicon at 4000 Mc/s.

- (a) $|Z_s| = 377\Omega$, $S = 0.1$ cm.
- (b) $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.

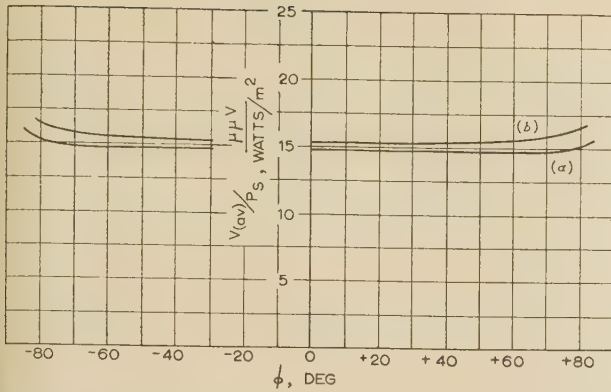


Fig. 10.—Hall e.m.f., $V_{(av)}$, for a given power, P_s , dissipated in loads of different impedances, $Z_s = |Z_s|/\phi$, for 500 ohm-cm n -type silicon at 4000 Mc/s.

- (a) $|Z_s| = 38.9\Omega$, $S = 0.1$ cm.
(b) $|Z_s| = 377\Omega$, $S = 0.1$ cm.

(5.3) Polythene

Assuming for polythene $\sigma/\omega\epsilon = 0.0005$ or $\sigma = 2.55 \times 10^{-4}$ mho/m with $\epsilon_r = 2.3$ and $\mu = \mu_0 = 4\pi \times 10^{-7}$ henry/m we find at 4000 Mc/s

$$\begin{aligned}\alpha &= 3.2 \times 10^{-2} \text{ nepers/m} \\ \beta &= 127 \text{ rad/m} \\ |\gamma| &= 127 \text{ per metre} \\ |Z| &= 248 \text{ ohms} \\ \delta &\approx 0^\circ\end{aligned}$$

An estimate of the value of \mathcal{H}_d is $5.8 \times 10^{-8} \text{ m}^3/\text{coulomb}$, and in spite of the small value of σ the mobility of the carriers responsible for the conductivity of this material must be so much smaller as to make \mathcal{H}_c negligible.

With polythene we are therefore concerned only with a displacement current contribution to the Hall e.m.f. and to the radiation pressure. In Fig. 11 curve (a) gives predicted values of this Hall e.m.f. When $Z_s = 377 \angle 67.5^\circ$ and $s = 0.1$ cm we find $V_{(av)} = V_{d(av)} = 2.7 \times 10^{-20} |E_s|^2$, so that for $|E_s| = 30 \text{ kV/cm}$, $V_{(av)} = 0.24 \mu\text{V}$, which might be just measurable in spite of the high impedance of the source. It is possible to increase $V_{(av)}$

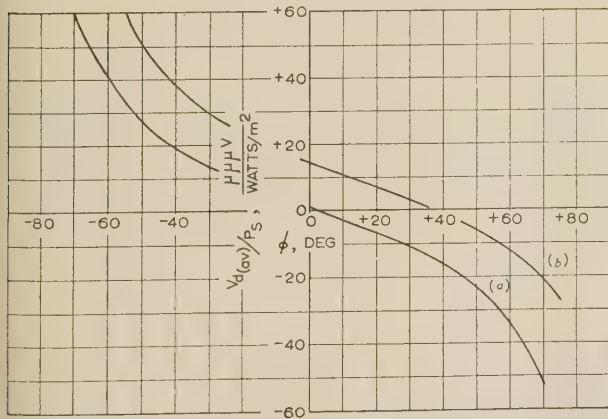


Fig. 11.—Hall e.m.f., $V_{d(av)}$, arising from the displacement component of the current for a given power, P_s , dissipated in a load of various phase angles, ϕ , at 4000 Mc/s.

- (a) Polythene: $\epsilon_r = 2.3$, $\sigma/\omega\epsilon = 0.0005$,
 $|Z_s| = 377\Omega$ and $S = 0.1$ cm.
(b) Germanium: 26 ohm-cm, n -type,
 $|Z_s| = 377\Omega$ and $S = 0.1$ cm.

to rather more than a microvolt by increasing s and making $2\beta s = 90^\circ$, but unfortunately this also increases the impedance of the source, so that no advantage may be gained. The corresponding displacement-current Hall effect for germanium is also shown in Fig. 11 for comparison.

As would be expected, the radiation pressure experienced by polythene is considerably smaller than for the semi-conductors in similar circumstances. The slab will tend to move in such a way as to make the stored energy in it a minimum, and consequently the pressure may be either positive or negative, as indicated in Fig. 12.

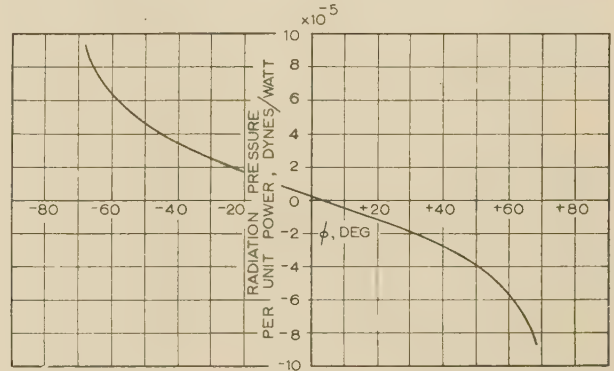


Fig. 12.—Radiation pressure, $p_{d(av)}$, arising from the displacement current for a given power, P_s , dissipated in a load of various phase angles, ϕ , for polythene at 4000 Mc/s.

$$|Z_s| = 377\Omega, S = 0.1 \text{ cm.} \\ \epsilon_r = 2.3, \sigma/\omega\epsilon = 0.0005.$$

(6) COMMENTS ON THE THEORY IN ITS APPLICATION TO MICROWAVE POWER MEASUREMENTS

All the calculations have been made for a plane wave. Larger values of Hall effect and radiation pressure at high frequencies could be obtained by suitable disposition of the slab when placed in a waveguide.

It is important to notice that the Hall e.m.f. does not depend upon the area of the surface presented to the incident wave, and there is nothing to be gained by increasing this area. Thus, with a semi-conductor a slab extending over a very small part of the wavefront is sufficient for the purpose and less likely to cause serious disturbance of the field distribution. On the other hand, the longitudinal dimension s is important in determining the Hall e.m.f. built up by this means. It is inherent in the transmission system that this e.m.f. is a measure of the power dissipated in the slab as well as of the power absorbed by the load.

The medium of the slab from which the Hall e.m.f. is derived must therefore be of high resistivity for accurate measurements of power transmitted through it. Of the semi-conductors, silicon can be much superior to germanium for this purpose. The effect of large losses in the slab is to flatten the curve of Hall e.m.f. against phase angle of the load, to lift it up above the horizontal axis and to make it asymmetrical (see Fig. 4). For a given $|Z_s|$ and $|Z|$ the value of $|\rho|$ from eqn. (28) is found to depend upon $\cos(\phi - \delta)$ in such a way that, since $0^\circ < \delta < 45^\circ$, the reflection coefficient must be larger for negative phase angles, ϕ , than for positive ones. Thus, with a larger proportion of the power thrown back when ϕ is negative, we find correspondingly smaller values of $V_{c(av)}$.

The alternative arrangement employing a resonant cavity containing a slab of semi-conductor and excited from the wave carrying the power⁸ is not nearly so exacting in the choice of a satisfactory medium for the slab. In general, for the transmission type of wattmeter s should be small so that $\alpha s \ll 1$, but

provided that this condition is satisfied, the larger the value of s the larger is the Hall e.m.f. obtained. Moreover, by suitable adjustment of βs , more favourable operating conditions can be established. Eqn. (33) for $V_{c(av)}$, representing the Hall e.m.f. arising from the conductivity of the medium, contains as a multiplier $[R_c\sigma/|\gamma|]$, and if b_c is the mobility of the current carriers in the medium, we have, from eqn. (14), $R_c\sigma = b_c/m$.

But $|\gamma| = \sqrt{(\omega\mu\sigma)[1 + (\omega\epsilon/\sigma)^2]}^{1/4}$

and if b_c can be considered constant, the value of $V_{c(av)}$, in so far as it depends upon this multiplier, would be expected to fall with increase of frequency. To achieve a large Hall e.m.f. when $\sigma \gg \omega\epsilon$ (such as is normally the case with germanium) we clearly require the material to have the highest possible ratio of mobility to conductivity. It is equally important to have a low conductivity from the aspect of the losses in the medium, and therein lies the particular advantage of silicon.

Radiation pressure on a semi-conductor like germanium differs in magnitude very little from the corresponding pressure on a metal surface. What the semi-conductor loses by the reduced conduction current it gains from displacement current, and the two components of pressure are, unlike the corresponding Hall e.m.f.'s, of the same order of magnitude. The expression for the radiation pressure exerted on a metal surface takes a much simpler form, namely

$$P_{(av)} = 2P_1\sqrt{(\mu_0\epsilon_0)}\left(1 - \sqrt{\frac{2\omega\epsilon_0}{\sigma}}\right)$$

and gives $P_{(av)}$ a little less than 1 dyne/watt. In all cases the radiation pressure for a uniform power density is directly proportional to the surface area presented to the incident wave.

In considering the contributions made by conductivity and displacement currents in n -type semi-conductors we dealt with these two factors as operating quite independently. The separate identity of mobile carriers in such material cannot, in fact, be maintained, and therefore it is possible that N_c and N_d are in some way interdependent, although this should not affect the validity of the argument given.

Eqn. (34) shows that, as one would expect, $V_{d(av)}$ increases with frequency and is zero when $\epsilon_r = 1$.

The close relationship between high-frequency Hall effect and radiation pressure offers the possibility of a valuable tool for further exploration of the basic characteristics of semi-conductors and dielectrics. If arrangements could be made to measure $V_{(av)}$ and $P_{(av)}$ at the same time, values of R_c and R_d would immediately emerge. In fact, the technique described by Cullen⁷ for the absolute calibration of a radiation-pressure wattmeter should be applicable to the equivalent Hall-effect instrument if provision is made for a small movement of the semi-conductor element, associating the mechanical forces on it with the Hall e.m.f. in it. These are matters for further research.

(7) CONCLUSIONS

The theory presented demonstrates clearly the close link between high-frequency Hall effect and radiation pressure. It predicts an equivalent Hall e.m.f. in a dielectric and suggests that observations based on this effect might offer a new approach in further exploration of the properties of such media. Attention is called to the possibility of simultaneous observations of Hall effect and radiation pressure on a semi-conductor element mounted in a waveguide as a means of establishing an absolute power-measuring instrument in which particular use is made of the Hall e.m.f.

The theory should also be of value in helping to elucidate the behaviour of calibrated microwave wattmeters employing the Hall effect in semi-conductors.⁸

(8) ACKNOWLEDGMENTS

The author is particularly indebted to Prof. Cullen for his earlier work on microwave radiation pressure and to his colleague Dr. John Brown, for most helpful discussions on this subject.

Mr. L. Stephenson, by his experimental work on the problem of power measurement by Hall effect in semi-conductors at 4000 Mc/s, has also provided a background of valuable information, which is gratefully acknowledged, as is also the support given to this work by the R. W. Paul Fund.

(9) REFERENCES

- (1) HALL, E. H.: 'On the New Action of Magnetism on Permanent Electric Current', *Philosophical Magazine*, 1880, **10**, p. 301.
- (2) BARLOW, H. E. M.: 'The Application of the Hall Effect in a Semi-Conductor to the Measurement of Power in an Electromagnetic Field' and 'The Design of Semi-Conductor Wattmeters for Power-Frequency and Audio-Frequency Applications', *Proceedings I.E.E.*, Papers No. 1654 M, June, 1954, and 1778 M, November, 1954 (**102 B**, pp. 17 and 186).
- (3) DONOVAN, B.: 'The Hall Effect in Metals at High Frequencies', *Proceedings of the Physical Society, A*, 1955, **68**, p. 1026.
- (4) CULLEN, A. L.: Discussion on Reference 2.
- (5) NICHOLS, E. F., and HULL, G. F.: 'The Pressure due to Radiation', *Proceedings of the American Academy of Arts and Science*, 1903, **38**, p. 559.
- (6) CULLEN, A. L.: 'Absolute Power Measurement at Microwave Frequencies', *Proceedings I.E.E.*, Monograph No. 23 M, February, 1952 (**99**, Part IV, p. 100).
- (7) CULLEN, A. L.: 'A General Method for the Absolute Measurement of Microwave Power', *ibid.*, Monograph No. 24 M, February, 1952 (**99**, Part IV, p. 112).
- (8) BARLOW, H. E. M., and STEPHENSON, L. M.: 'The Hall Effect and its Application to Power Measurement at Microwave Frequencies', *ibid.*, Paper No. 1913 R, January, 1955 (**103 B**, p. 110).

RADIO PROPAGATION OVER A DISCONTINUITY IN THE EARTH'S ELECTRICAL PROPERTIES—I

By T. B. A. SENIOR, M.Sc., Ph.D.

(The paper was first received 20th May, 1955, and in revised form 17th May, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

The problem considered is that in which vertically polarized waves are propagated across a single straight-line discontinuity in the electrical properties of a smooth flat earth, the discontinuity being such as might occur at a (rather idealized) coastline. It is further specialized by the assumption that one of the media (e.g. sea) has infinite conductivity, and this is replaced by an infinitely thin, perfectly conducting half-plane lying in the interface of the air and the land medium, the latter now being taken to fill completely the region below the interface.

With this model a method of solution is proposed which is a union of ray theory and rigorous diffraction theory. The presence of the land is accounted for by the introduction of a suitable image field and the problem thereby reduced to one in which two fields are incident upon a perfectly conducting sheet situated in free space. The technique is illustrated by application to the case of a three-dimensional plane wave, and this solution forms the basis for the later work.

When the transmitter is at a finite distance from the coastline the solution can be obtained by expressing the incident field as an angular spectrum of plane waves the propagation of which has already been studied. This procedure is first applied to the two-dimensional problem of a line source and the results are shown to be in close agreement with previous work on the subject. On examination, however, the solution is found to be made up of the field of the actual transmitter diffracted at the coastline, together with the diffracted field arising from an image transmitter of strength equal to the Fresnel reflection coefficient appropriate to a particular angle of incidence, and this suggests an even more elementary way of tackling the problem. For any type of transmitter all that is now required is the solution of the corresponding Sommerfeld diffraction problem, and with this new approach the case of a point source is treated. The result is almost identical with that found by resolution into plane waves. Moreover, a comparison with the solution for a line source shows that the factor by which the free-space field must be multiplied to give the total field is of the same form for both transmitters, the factor for a point source being obtainable from that for a line source by replacing each two-dimensional distance by the corresponding three-dimensional one. In particular this justifies the application to each radial separately of the formulae for normal incidence on the coastline and enables the analytical difficulties associated with three-dimensional propagation to be avoided.

The nature of the field is then studied with particular reference to the rapid changes of intensity and phase just beyond the coastline; the former is generally termed the 'recovery effect', whilst the latter gives rise to the phenomenon of coastal refraction. The use of height-gain factors is considered and it is shown that, for sufficiently small elevations of transmitter and receiver, the factors appropriate to homogeneous earths may be applied. Several numerical examples are given.

(1) INTRODUCTION

The theory of ground-wave propagation over a homogeneous smooth earth has been given adequate mathematical treatment by a number of workers, and as a result the solutions which have been obtained are, in general, universally accepted. In practice,

however, the earth's crust is far from homogeneous and it is frequently essential that this fact be taken into account. Even if the only inhomogeneities to be admitted are variations in the electrical properties of an earth which is otherwise smooth and flat, the problem remaining is still one of great complexity, and not until recent times have theoretical attacks upon it met with any success. All have been under certain idealizing assumptions, two extreme cases being capable of treatment:

(a) One is that in which the changes in earth constants are irregular and ill-defined; some average value of the constants can then be inserted into the solution appropriate to a homogeneous earth, or alternatively a purely statistical theory can be developed (e.g. see Reference 1);

(b) The other is that in which there is a single boundary, such as a coastline, dividing the earth into two regions each of which is homogeneous.

It is this last case which will be considered here, and the problem is therefore the determination of the field strength and phase over a smooth flat earth possessing a single discontinuity in electrical properties.

Several empirical methods have been proposed to enable the field strength to be deduced from the attenuation curves for a homogeneous earth. The most important of these is due to Millington,² and has some theoretical backing in that it is designed to satisfy the reciprocity condition and to take into account the 'geometric mean' formula of Eckersley.³ It yields results in good agreement with experiment over a wide range of frequencies and predicts the rather startling changes in field strength which can occur just beyond a boundary, whilst more recent work⁴ suggests that it may also be applicable to phase.

The first successful attempt to provide a full analytic solution of the problem was made by Grünberg^{5,6} and was aimed primarily at a study of the deviation of a plane wave incident on a coastline. By using approximate boundary conditions at the air-land interface (the sea was assumed perfectly conducting) an integral equation was obtained from which it was established that the original direction of propagation is restored at large distances from the coast.

Grünberg's work was generalized by Feinberg in a series of papers of which the third⁷ treats this topic, but with the added complication of a transmitter at a finite distance. The analysis is so manipulated that an assumed value may reasonably be substituted for the unknown field component under the integral sign, and the problem is thus reduced to one of integration. A related treatment has been given by Furutsu,⁸ who derived an inhomogeneous integral equation for the single-component Hertz vector representing the total field of a vertical dipole in the presence of the two media. The equation is solved by successive substitution of the solutions for earths composed of each medium separately, and the final solution then appears as a series of terms involving multiple integrals over known functions, only the first few terms being evaluated. Aono and Muramatsu⁹ have provided experimental confirmation of the results.

The solutions produced by this last method are analogous to those of Clemmow,¹⁰ who considered the two-dimensional case

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Senior is at the Radar Research Establishment, and was formerly at Cambridge University.

of a line-source transmitter parallel to the coastline. Although the problem is thereby limited in scope, it now becomes capable of exact solution in closed form. In the first part of the paper one medium is taken to be perfectly conducting. The problem is initially solved for an incident plane wave by expressing the scattered field as an angular spectrum, and this leads to a pair of dual integral equations which are handled using complex variable techniques. For a line source the solution is obtained by integrating that for a plane wave. In the second part of the paper the restriction that one medium be perfectly conducting is waived, and an integral equation formulation then follows by adopting approximate boundary conditions at the earth's surface.

With all the above methods the analysis is both extensive and complicated, even though the solutions themselves reduce to quite simple forms in many of the cases of practical interest. This fact suggests that an alternative approach might be possible whereby the sacrifice of a little rigour would allow the most important results to be derived in a manner which would enable the underlying physical mechanism to be more clearly understood. The present paper is directed towards that goal, and it is believed that the simplicity attained compensates for any loss of accuracy.

An idealized type of discontinuity is assumed in which a straight-line boundary separates two media, one of which is perfectly conducting. The incident field is a plane or cylindrical wave, or a dipole field, and whilst the first of these is merely a particular case of the other two and obtainable from them by taking the transmitter off to infinity, it is in many ways the most fundamental. It is treated in Section 3 and serves to illustrate the method and to indicate its limitations. The solution for a transmitter at a finite distance can be found by integrating that for a plane wave with respect to its angle of incidence: this is done for a line source in Section 4 and a comparison with Clemmow's exact result¹⁰ confirms the adequacy of the present technique. When the expression for the field is examined, however, its form is seen to imply that any transmitter can be dealt with by a process entirely equivalent to that for a plane wave and employing only the solution of the appropriate Sommerfeld diffraction problem. This is the crux of the method and is described in Section 4.4. The application to a point source is carried out in Section 5 and provides results of considerable practical significance. Moreover, oblique incidence on the boundary can now be considered, leading to a study of coastal refraction, but the interpretation of the solution in this light is reserved for a subsequent paper.

(2) THE MODEL FOR THE IDEALIZED PROBLEM

The problem considered is that in which vertically polarized waves are propagated across a straight-line discontinuity in the electrical properties of a smooth flat earth. The situation is further specialized by the assumption that one of the media (medium 2) has infinite conductivity and this is replaced by an infinitely thin, perfectly conducting half-plane lying in the interface of the air (regarded as free space) and medium 1, the latter being taken to fill completely the region below the interface (see Fig. 1). Such a model, of course, is a valid representation of the physical problem only as regards the determination of the field above the earth's surface and attention will be confined to this region.

The assumption of perfect conductivity for an earth constituent may sometimes be justified, sea water, for example, often fulfilling the condition to an adequate degree of accuracy. For this reason media 1 and 2 will be referred to as land and sea respectively, but whilst consideration will be given primarily to an investigation of the behaviour of the field near to a coastline, the solutions obtained are appropriate to any mixed-path problem in which the above assumptions may be supposed satisfied.

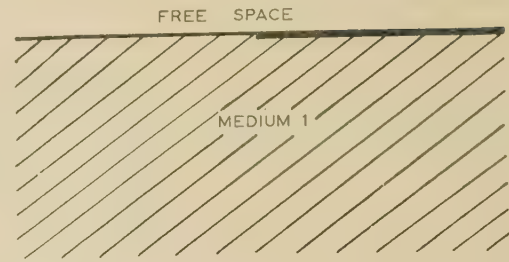


Fig. 1.—The model of the idealized problem.

Furthermore, under most practical conditions the radiation penetrates negligibly into the ground, so that the results provided by the model of Fig. 1 are not likely to be significantly different from those produced by a more realistic model in which medium 2 has a finite depth. Indeed Höller¹¹ has attempted to assess the effect of the boundary condition at the land-sea interface and has concluded that it may be neglected.

For the purposes of the analysis a necessary feature of the problem is the straightness of the coastline, but in any practical application of the solution this restriction may be relaxed. The work of Feinberg¹² indicates that the nature of the propagation is almost entirely determined by the properties of that portion of the earth contained by an ellipse having foci at the transmitter and receiver. The ellipse has axes $1/\gamma(\gamma d + 1)$ and $1/\gamma\sqrt{(2\gamma d + 1)}$, where d is the distance from the transmitter to the receiver and γ is the propagation coefficient, and only within this region is it essential to regard the coastline as straight.

(3) THE SOLUTION FOR AN INCIDENT PLANE WAVE

The approximate method is here described and illustrated by application to the case of a three-dimensional plane wave incident upon the model (Section 3.2). An examination of the degree to which this solution satisfies the conditions of the problem then enables us to assess the accuracy of the technique (Section 3.3).

(3.1) The Method

An appreciation of the method can best be obtained from a study of the solution for the diffraction of a plane wave by a perfectly conducting half-plane in free space.

In terms of Cartesian co-ordinates (x, y, z) the perfectly conducting sheet occupies the half-plane $y = 0, x > 0$; the edge of the sheet is coincident with the z -axis and will later represent the coastline. Polar co-ordinates (r, θ) are also introduced and are such that $x = r \cos \theta, y = r \sin \theta$.

If the incident field (denoted by superscript i) is a two-dimensional H -polarized wave (i.e. one whose only component of magnetic vector is parallel to the z -axis) and if*

$$H_z^i = \epsilon^{j\gamma r \cos(\theta - \alpha)} \quad (1)$$

where α is the angle of incidence, then

$$H_z = \frac{\epsilon^{j\pi/4}}{\sqrt{\pi}} \left\{ \epsilon^{j\gamma r \cos(\theta - \alpha)} F[-\sqrt{(2\gamma r)} \cos \frac{1}{2}(\theta - \alpha)] + \epsilon^{j\gamma r \cos(\theta + \alpha)} F[-\sqrt{(2\gamma r)} \cos \frac{1}{2}(\theta + \alpha)] \right\}$$

(see, for example, Clemmow¹³), and the other non-zero field components are given by

$$E_x = -\frac{jZ}{\gamma} \frac{\partial H_z}{\partial y}, \quad E_y = \frac{jZ}{\gamma} \frac{\partial H_z}{\partial x}$$

* Rationalized M.K.S. units are used and a time factor $e^{j\gamma ct}$ is suppressed throughout.

where $Z = 1/Y$ is the intrinsic impedance of free space. In the above equation $F(a)$ denotes the Fresnel integral defined by

$$F(a) = \int_a^{\infty} e^{-js^2} ds$$

On the surface $\theta = \pi$ (the half-plane complementary to the sheet), a consequence of the symmetry of the scattered field is that its vertical electric and tangential magnetic components vanish. The remaining component is

$$E_x = E_x^i - 2 \frac{e^{j\pi/4}}{\sqrt{\pi}} e^{-j\gamma r \cos \alpha} F[\sqrt{(2\gamma r)} \sin \frac{1}{2}\alpha]$$

and if $\sqrt{(2\gamma r)} \sin \frac{1}{2}\alpha$ is sufficiently large to justify replacing the Fresnel integral by its asymptotic expansion,

$$E_x - E_x^i \sim \frac{e^{-j\gamma r - j\pi/4}}{\sqrt{(2\gamma r)} \sin \frac{1}{2}\alpha}$$

which tends to zero with increasing distance from the diffracting edge. The 'backwash' is therefore small, and with the exception of a strip adjacent to the edge, the field on the surface $\theta = \pi$ is sensibly unaffected by the presence of the perfectly conducting sheet. Moreover, the width of the strip within which the influence of the sheet cannot be ignored is a function of the angle α at which the field is incident and is least when $\alpha = \pi$.

Similar conclusions hold when the incident field is E -polarized.

We now turn our attention to the model and assume a two-dimensional plane wave incident upon it from the land at an angle α not differing greatly from π . It is convenient to regard this as a boundary-value problem, the object being the determination of the field in the region above the earth's surface. At the air-land interface the boundary condition to be satisfied is that corresponding to a land of infinite extent; at the air-sea interface the condition is that for a perfect conductor. If the land alone were present, the field above it would consist of the original incident wave, together with a reflected wave multiplied by the Fresnel reflection coefficient ρ appropriate to the angle at which the field is incident upon the earth, so that the effect of the land may be accounted for by the introduction of a suitable image field. Also, the field scattered by a perfectly conducting sheet in free space is relatively small at points on the surface complementary to the sheet, the field there being essentially that which is incident. These results suggest the following procedure for treating the model:

(a) The boundary condition at the air-land interface is satisfied (approximately) by introducing the image field appropriate to a land of infinite extent: the presence of the land may then be ignored.

(b) The effect of the perfectly conducting sheet upon the (two) plane waves is determined: the solution of this diffraction problem is easily found.

The method may be represented symbolically as follows. Let $F^i(\alpha)$ be a typical vector in the incident field and $F^s(\alpha)$ the corresponding vector in the field scattered by the perfectly conducting sheet *when situated in free space*. As a result of the operation (a), the problem is reduced to that of finding the effect of the sheet upon the modified incident field

$$F^i(\alpha) + \rho F^i(2\pi - \alpha)$$

and the total field is therefore

$$F^i(\alpha) + F^s(\alpha) + \rho[F^i(2\pi - \alpha) + F^s(2\pi - \alpha)] \quad (2)$$

It will be seen that the method is a union of ray theory and rigorous diffraction theory. The use of ray theory is, of course, only valid under certain circumstances; in particular, it requires that the field be incident from the direction of the land, so making possible the imaging process.

(3.2) Application to a Three-Dimensional Plane Wave

Although this can be regarded as merely an illustration of the above method, the solution is of theoretical importance in that any field may be expressed as an angular spectrum of three-dimensional plane waves.

It is convenient to choose an incident field having no component of the magnetic vector perpendicular to the earth. The field is then directly analogous to that for a vertical electric dipole (from which it may be obtained by allowing the dipole to recede to infinity), and, moreover, the imaging requires the introduction of only one Fresnel reflection coefficient. Such a plane wave may be described in terms of the function

$$W = e^{j\gamma(x \cos \alpha \cos \beta + y \sin \alpha \cos \beta - z \sin \beta)}$$

where $\pi/2 - \beta$ is the angle between the projection of the wave normal on the earth's surface and the coastline, and the corresponding field components are given by eqn. (29).

In the solution of the problem by the method of Section 3.1 the boundary condition at the air-land interface is satisfied by combining the incident field with that obtained by imaging in the land. Since the electric vector is entirely in the plane of incidence, the Fresnel reflection coefficient is given approximately by

$$\rho = \frac{\sin \chi - \eta}{\sin \chi + \eta}$$

where η is the reciprocal of the complex refractive index of the land (relative to free space) and χ is the angle which the wave normal makes with its projection on the surface of the earth. In terms of the angles α and β ,

$$\sin \chi = \cos \beta [\sin^2 \alpha + (1 + \cos \alpha)^2 \sin^2 \beta]^{1/2}$$

and since $(\pi - \alpha)$ is small in all cases of interest, it is reasonable to neglect $(1 + \cos \alpha)^2 \sin^2 \beta$ by comparison with $\sin^2 \alpha$ to give

$$\rho = \frac{\sin \alpha \cos \beta - \eta}{\sin \alpha \cos \beta + \eta} \quad (3)$$

The remaining problem is one of pure diffraction and may be treated as in Section 11, the resulting field being of the form

$$(E, H) = [E(\alpha), H(\alpha)] + \rho[E(2\pi - \alpha), H(2\pi - \alpha)] \quad (4)$$

The components of $E(\alpha)$, $H(\alpha)$ are given by eqns. (30), and from these $E(2\pi - \alpha)$, $H(2\pi - \alpha)$ may be deduced by appropriate change of α .

(3.3) The Fulfilment of the Boundary Conditions

Although the solution obtained by this method is not exact, it does fulfil most of the conditions of the problem. Those at infinity and at the air-sea interface are automatically satisfied, as is the condition upon the order of singularity at the coastline itself. It is only left to consider the air-land interface, and for this purpose it is sufficient to take the case of an incident three-dimensional plane wave.

The refractive index $1/\eta$ of the land is (by assumption) large, and this justifies the application of the approximate boundary conditions

$$E_z = -\eta Z H_x, \quad E_x = \eta Z H_z$$

on the surface $\theta = \pi$. The postulated solution fails to satisfy these completely, and if the amounts by which it does so are written

$$I = E_z + \eta Z H_x, \quad J = E_x - \eta Z H_z$$

then, by inserting the expressions for the field components,

$$I \simeq 4\eta e^{-j\pi/4} \frac{\cos \frac{1}{2}\alpha \sin \beta \cos^2 \beta}{\eta + \sin \alpha \cos \beta} \frac{e^{-j\gamma(r \cos \beta + z \sin \beta)}}{\sqrt{(2\pi\gamma r \cos \beta)}}$$

$$J \simeq 4\eta e^{-j\pi/4} \frac{\cos \frac{1}{2}\alpha \sin^2 \beta \cos \beta}{\eta + \sin \alpha \cos \beta} \frac{e^{-j\gamma(r \cos \beta + z \sin \beta)}}{\sqrt{(2\pi\gamma r \cos \beta)}}$$

The Fresnel integral has here been replaced by the first term of its asymptotic expansion. A study of the integral using tabulated values (e.g. Rankin¹⁴) shows that negligible error is incurred in this process if

$$\sqrt{(2\gamma r \cos \beta)} \sin \frac{1}{2}\alpha > 5$$

and since $\sin \frac{1}{2}\alpha$ is effectively unity, this is equivalent to

$$r \cos \beta > 2\lambda$$

where λ is the wavelength.

With this restriction upon r and β , I and J are almost identical, both decreasing in magnitude with increasing distance from the coastline, whilst the fact that $(\pi - \alpha)$ is small implies that I and J are of the second order of smallness by virtue of their dependence on α . The postulated solution can therefore be regarded as satisfying the boundary conditions over that part of the air-land interface for which $r \cos \beta > 2\lambda$. The excluded region is a strip of width $2\lambda \sec \beta$ and here the conditions are violated to an increasing extent as r decreases.

It will be observed that the obliquity β has relatively little effect upon these conclusions; in particular, it affects the width of the region wherein the boundary conditions are unfulfilled only through a factor $\sec \beta$. The case $\beta = 0$, however, requires special consideration. I is then identically zero, so that the first condition is satisfied exactly. But this is not true of the second; although J vanishes to the order $r^{-1/2}$, it has a non-zero term of the order $r^{-3/2}$ (as may be seen by including the second term in the asymptotic expansion of the Fresnel integral), and this serves to emphasize the necessity of using *both* boundary conditions at the air-land interface when assessing the accuracy of the solution.

(4) THE SOLUTION FOR A LINE SOURCE

A natural extension of the above method is to the case of a transmitter at a finite distance from the coastline. The problem is here considered in its two-dimensional form, the transmitter being a line source which, in free space, would propagate the cylindrical wave

$$H_z^i = \sqrt{\left(\frac{\pi}{2}\right)} e^{-j\pi/4} H_0^{(2)}(\gamma R) \sim \frac{e^{-j\gamma R}}{\sqrt{(\gamma R)}}$$

where $H_0^{(2)}(\gamma R)$ is the zero-order Hankel function of the second kind and R is the distance from the source. The co-ordinates of the source are taken to be (r_0, θ_0) , with θ_0 not very different from π .

(4.1) The Analysis

By introducing the appropriate phase factor, Sommerfeld's integral representation of the Hankel function becomes

$$H_0^{(2)}(\gamma R) = \frac{1}{\pi} \int_{S(\pi/2)} \exp[-j\gamma r_0 \cos(\theta_0 - \alpha)] \exp[j\gamma r \cos(\theta - \alpha)] d\alpha$$

where $S(\pi/2)$ is the steepest descent path through the angle $\pi/2$ (see Fig. 2). This enables the incident field to be expressed as an angular spectrum of plane waves whose propagation has already been studied. If the solution for the two-dimensional plane wave [eqn. (1)] is multiplied by

$$1/\sqrt{(2\pi)} \exp(-j\pi/4) \exp[-j\gamma r_0 \cos(\theta_0 - \alpha)]$$

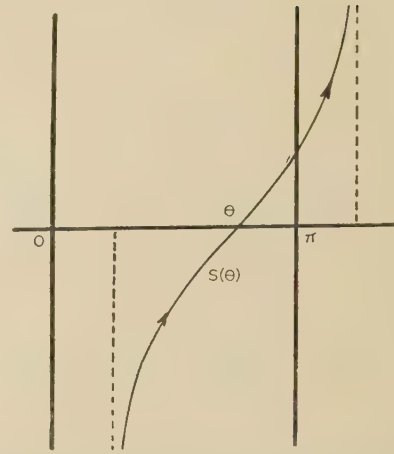


Fig. 2.—The steepest descent path through the angle θ .

and integrated with respect to α , the total field produced by the line source in the presence of the model can be broken up into geometrical optics and diffraction terms and is

$$H_z = H_z^g + H_z^d \quad (5)$$

where

$$H_z^g = \frac{\exp(-j\pi/4)}{\sqrt{(2\pi)}} \int_{S(\pi/2)} \left\{ \exp[j\gamma r \cos(\theta - \alpha)] + (1, \alpha)_{\theta_0} \exp[j\gamma r \cos(\theta + \alpha)] \right\} \exp[-j\gamma r_0 \cos(\theta_0 - \alpha)] d\alpha \quad (6)$$

and

$$H_z^d = \frac{\exp(j\pi/4)}{\pi\sqrt{(2\pi)}} \int_{S(0)} \int_{S(0)} [1 - \rho(\sin \alpha + \theta_0)] \frac{\cos \frac{1}{2}(\alpha + \theta_0) \cos \frac{1}{2}(\phi + \theta)}{\cos(\alpha + \theta_0) + \cos(\phi + \theta)} \exp[-j\gamma(r_0 \cos \alpha + r \cos \phi)] d\alpha d\phi \quad (7)$$

The symbol $(1, \alpha)_{\theta_0}$ in eqn. (6) is defined by

$$(1, \alpha)_{\delta} = \begin{cases} 1 & \text{for } 0 \leq \theta < \pi - \delta \\ \rho(\sin \alpha) & \text{for } \pi - \delta < \theta \leq \pi \end{cases}$$

the suffix referring to the division in the range of θ .

In view of the approximate nature of the constituent plane-wave solution, it would be superfluous to seek an exact evaluation of these integrals, although this would be possible using the formula for $\rho(\sin \alpha)$ [eqn. (3) with $\beta = 0$]. An alternative treatment of sufficient accuracy follows from a consideration of

$$\int_{\rho S(\pi/2)} \rho(\sin \alpha) \exp[-j\gamma r_0 \cos(\theta_0 - \alpha)] \exp[j\gamma r \cos(\theta + \alpha)] d\alpha$$

which may be written as

$$\int_{S(\pi/2)} \rho(\sin \alpha) \exp[j\gamma S \cos(\alpha + \psi)] d\alpha$$

where $S \cos \psi = x - x_0$ and $S \sin \psi = y + y_0$, S being the distance from the image transmitter. The only singularities of $\rho(\sin \alpha)$ are poles at those values of α for which $\sin \alpha = -\eta = -\sin \alpha_B$ (α_B is the Brewster angle of the earth medium), and since all of these lie outside the strip $0 < R_e \alpha < \pi$, the path of integration can be displaced to $S(\pi - \psi)$. It is now possible to put $\alpha = \pi - \psi$ in that part of the integrand which is slowly

varying in the vicinity of the saddle point, and this is true of $\rho(\sin \alpha)$, provided that $|\psi + \alpha_B|$ and $|\pi - \psi + \alpha_B|$ are not small. For a transmitter over land only the former requirement affects us, and if this is satisfied

$$\int_{\phi S(\pi/2)} (\sin \alpha) \exp [j\gamma S \cos (\alpha + \psi)] d\alpha = \pi \rho(\sin \psi) H_0^{(2)}(\gamma S) \\ \sim \sqrt{(2\pi)} \exp (j\pi/4) \rho(\sin \psi) \frac{\exp (-j\gamma S)}{\sqrt{(\gamma S)}}$$

The diffraction field H_z^d may be treated in a similar manner by removing the factor $[1 - \rho(\sin \alpha + \theta_0)]$ from the integral at the predominant value $\alpha = 0$. The remaining integral can be reduced¹⁵ to a sum of Fresnel integrals, leading to an expression for the total field in the form

$$H_z = \left[1 + (1, \psi)_{\theta_0} \exp [-j\gamma(S - R)] \right. \\ \left. - \frac{\exp (j\pi/4)}{\sqrt{\pi}} [1 - \rho(\sin \theta_0)] \{ F[\gamma(R_1 - R)]^{1/2} \right. \\ \left. \pm \exp [-j\gamma(S - R)] F[\gamma(R_1 - S)]^{1/2} \} \right] \frac{\exp (-j\gamma R)}{\sqrt{(\gamma R)}} \quad (8)$$

where S has been replaced by R in those factors which are relatively unaffected by small changes in γ and y_0 , and $R_1 = r + r_0$. The upper or lower sign is chosen according as $\theta < \pi - \theta_0$ or $\theta > \pi - \theta_0$ respectively, and across the shadow edge $\theta = \pi - \theta_0$ the field is continuous.

(4.2) The Nature of the Approximations

The basis of the method is the integration of the solution for a plane wave with respect to its angle of incidence α . Although this solution holds good only if α does not differ greatly from π , the integration is carried over all angles of incidence (real and complex), and the process therefore requires careful examination in order to establish its validity.

Taking first the expression for H_z^g [eqn. (6)], the main contribution to the integral is provided by those values of α in the immediate vicinity of the saddle point $\alpha = \pi - \psi$, the effect of plane waves incident at other angles being negligible by comparison. The saddle point lies within the range of α for which the plane-wave solution is valid if ψ is small, and this is satisfied if $\gamma + y_0 \ll R$. With this limitation upon the allowed positions of transmitter and receiver, the integration is justified. A similar result applies to the expression for H_z^d [eqn. (7)]; here the saddle point is $\alpha = \theta_0$ and lies within the required range of α if the angle of elevation of the transmitter is small.

In the evaluation of H_z^g and H_z^d it was assumed that the Fresnel reflection coefficient could be regarded as a polar diagram—a procedure which is equivalent to a partial use of ray theory and is valid if $|\psi + \alpha_B|$ is not too small. It is difficult to frame a precise statement of this condition, but a rough criterion may be deduced from homogeneous earth analysis and is $\gamma(\gamma + y_0)|\sin \alpha_B| \gg 1$.

The restrictions implied by the method can therefore be summarized as

$$x_0 < 0 \quad (9a)$$

$$\gamma R, \gamma S, \gamma r, \gamma r_0 \gg 1 \quad (9b)$$

$$R \gg \gamma + y_0, \quad r_0 \gg y_0 \quad (9c)$$

$$\gamma(\gamma + y_0)|\sin \alpha_B| \gg 1 \quad (9d)$$

and would be fulfilled under a fairly wide variety of practical conditions. For the success of the method, however, it is

necessary that the transmitter be over land and at a large distance from the coastline. The corresponding problem of a transmitter over the sea cannot be treated in itself, but the solution may be deduced from the above by the reciprocity condition concerning the interchangeability of transmitter and receiver.

(4.3) The Receiver at Ground Level

Although the solution is not directly applicable to a study of the ground-to-ground field, it is possible to take either the transmitter or the receiver on the earth's surface providing the other is elevated in accordance with the inequality (9d). The fact that S equals R then simplifies eqn. (8) considerably.

A further reduction can be achieved in the particular case of a ground receiver which is not far from the coast. Since

$$R^2 = R_1^2 - 2rr_0(1 + \cos \theta_0)$$

it follows that, for small values of $\pi - \theta_0$,

$$\sqrt{[\gamma(R_1 - R)]} \approx y_0 \sqrt{\left(\frac{\gamma r}{2Rr_0} \right)}$$

which is itself small if R is not very different from r_0 . The Fresnel integral can thus be replaced by its series expansion

$$F(a) = \frac{1}{2}\sqrt{(\pi)}\varepsilon^{-j\pi/4} - a + 0(a^3)$$

to give

$$H_z = \frac{2y_0/R}{y_0/R + \eta} \frac{\varepsilon^{-j\gamma R}}{\sqrt{(\gamma R)}} \quad (10)$$

or

$$H_z = \frac{2y_0/r_0}{y_0/r_0 + \eta} \left[1 + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma R r}{r_0} \right)} \right] \frac{\varepsilon^{-j\gamma R}}{\sqrt{(\gamma R)}} \quad (11)$$

for a receiver on the surface of the land or sea respectively.

It is of interest to compare this result with Clemmow's exact solution.¹⁰ His treatment leads naturally to an expression for the ground-to-ground field from which the field for a transmitter at a height y_0 may be deduced using the concept of height-gain factor is $(1 + j\eta\gamma y_0)$, and if it is assumed that $|\eta\gamma y_0| \gg 1$, the field differs from that of eqns. (10) and (11) merely in having η in place of $y_0/R + \eta$ and $y_0/r_0 + \eta$. This is a valid approximation if $R \approx r_0$ and $y_0 \ll r_0$, and, indeed, both these restrictions were imposed in the derivation of eqns. (10) and (11).

It should be noted that a comparison has been effected only between simplified versions of the formulae produced by the two methods. Whilst this does not provide full confirmation of the present technique, the general agreement between the formulae, and also between the conditions under which they are valid, is encouraging. The sole distinction relates to the conditions upon R and r_0 , the reduction of the exact expression for H_z requiring that the numerical distances

$$|\eta\sqrt{(\gamma R/2)}|, \quad |\eta\sqrt{(\gamma r_0/2)}|, \quad |\eta\gamma y_0|$$

rather than the distances themselves, be large compared with unity.

(4.4) A Simplified Approach

The solution for a line source has been obtained by integration of that for a plane wave, and the complications of analysis are largely due to the presence of the Fresnel reflection coefficient in the integrand. To the desired accuracy, however, the factor $\rho(\sin \alpha)$ may be removed from the integrand at the predominant value of α —a procedure equivalent to regarding the effect of the land as being catered for by the introduction of an image transmitter of suitable strength. A more elementary approach to the problem now suggests itself.

If eqn. (8) is examined, it is seen that the expression for the total field divides naturally into two parts. The first of these is merely the free-space field of the transmitter diffracted at the coastline, whilst the second is a close approximation to the field, similarly diffracted, of an image transmitter of strength $\rho(\sin \theta_0)$. There is a discrepancy in the geometrical optics terms, a factor $\rho(\sin \theta_0)$ having replaced $\rho(\sin \psi)$, but the difference is zero on the shadow edge and negligible for small elevations of transmitter and receiver; indeed, the conditions (9c) demand a restriction to such elevations. It follows that, in the analysis of the field into plane waves, each of these may be taken to have the same Fresnel reflection coefficient. The solution is then analogous to that for the limiting case of a plane wave and is given by eqn. (4) with θ_0 instead of α .

The interpretation of the solution in this light leads to a considerable simplification in the treatment of propagation from a transmitter at a finite distance from the coastline. For any type of transmitter all that is needed is the solution of the corresponding Sommerfeld problem, and this technique will now be applied to the case of a point source.

(5) THE SOLUTION FOR A POINT SOURCE

A transmitter of some practical importance is that represented mathematically by an electric dipole with axis normal to the earth's surface. The free-space field is then vertically polarized and is given by

$$E = \left(\frac{\partial^2 \Pi}{\partial x \partial y}, \frac{\partial^2 \Pi}{\partial y^2} + \gamma^2 \Pi, \frac{\partial^2 \Pi}{\partial y \partial z} \right), \quad H = j\gamma Y \left(-\frac{\partial \Pi}{\partial z}, 0, \frac{\partial \Pi}{\partial x} \right) \quad (12)$$

where $\Pi = \varepsilon^{-j\gamma R}/\gamma R$, R being the distance from the dipole, the co-ordinates of which are (x_0, y_0, z_0) ; a polar diagram factor may be included if desired, but for convenience this will be omitted. As in Section 4, it is again assumed that the transmitter is over land and that its angle of elevation is small.

In order to apply the method of Section 4.4 the solution is required for the problem of the field [eqn. (12)] incident upon a perfectly conducting half-plane in free space. This has been obtained by Senior,¹⁶ and if attention is confined merely to dominant terms

$$E_y = \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} [F(-\tau_R) + \varepsilon^{-j\gamma(S-R)} F(-\tau_S)] E_y^i \quad (13)$$

where

$$\tau_R = 2\sqrt{\left(\frac{\gamma r r_0}{R_1 + R}\right)} \cos \frac{1}{2}(\theta - \theta_0)$$

$$\tau_S = 2\sqrt{\left(\frac{\gamma r r_0}{R_1 + S}\right)} \cos \frac{1}{2}(\theta + \theta_0)$$

with

$$R_1 = r + r_0 \text{ and } S = [(x - x_0)^2 + (y + y_0)^2 + (z - z_0)^2]^{1/2}$$

An approximate solution of the propagation problem now follows by adding to eqn. (13) the corresponding field of an image source of strength

$$\rho = \frac{\sin \theta_0 \cos \Psi - \eta}{\sin \theta_0 \cos \Psi + \eta}$$

where $\tan \Psi = (z - z_0)/R_1$, ρ being the Fresnel reflection coefficient of the land. The expression for this auxiliary field may be deduced from eqn. (13) by changing θ_0 into $2\pi - \theta_0$ (a transformation which also implies $R \leftrightarrow S$, together with a change in sign of y_0), and the total field is therefore

$$E_y = \left\{ 1 + \varepsilon^{-j\gamma(S-R)} - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} (1 - \rho) [F(\tau_R) + \varepsilon^{-j\gamma(S-R)} F(\tau_S)] \right\} E_y^i \quad (14)$$

If the incident field had been expressed as an angular spectrum of plane waves, as in Section 4.1, a solution would have been found which differs from eqn. (14) only in the second term on the right-hand side, and under most practical circumstances this discrepancy is negligible (cf. Section 4.4). In view of this agreement it is possible to state the restrictions upon the validity of eqn. (14) by reference to the analysis involved in the derivation using an angular spectrum. The conditions turn out to be almost identical with those of Section 4.2, and, in fact, a study of the approximations shows that it is only necessary to add to the inequalities (9) the requirement that the obliquity Ψ be not too great (say, $\Psi < \pi/3$), a limitation which is implicit in the fundamental solution for a plane wave.

An additional feature of eqn. (14) may be seen by casting it into a slightly different form. From the definition of τ_R it follows that

$$\tau_R^2 = \gamma(R_1 - R) + \frac{\gamma}{R_1 + R}(z - z_0)^2$$

and if the obliquity Ψ is not too large,

$$\tau_R \approx \sqrt{[\gamma(R'_1 - R)]}$$

where $R'_1 = r' + r'_0$, the primes denoting distances relative to an origin of co-ordinates at the point where the direct ray from transmitter to receiver crosses the coast and measured radially from the transmitter. Similarly

$$\tau_S \approx \pm \sqrt{[\gamma(R'_1 - S)]} \quad \pm \text{for } \theta \gtrless \tau - \theta$$

and a reasonable approximation to the expression for E_y is therefore

$$E_y = \left\{ 1 + \varepsilon^{-j\gamma(S-R)} - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} (1 - \rho) \left[F[\gamma(R'_1 - R)]^{1/2} + \varepsilon^{-j\gamma(S-R)} F\left\{ (\pm [\gamma(R'_1 - S)]^{1/2}) \right\} \right] \right\} E_y^i \quad (15)$$

This should be compared with the solution for a line source [eqn. (8)]. Since

$$F(a) + F(-a) = \sqrt{(\pi)} \varepsilon^{-j\pi/4}$$

the factor by which the free-space field must be multiplied to give the total field is of the same form for both types of transmitter, the factor for a point source being obtainable from that for a line source by replacing each two-dimensional distance by the corresponding three-dimensional one; whilst this was known to be true for a homogeneous earth (Booker and Clemmow¹⁷), its validity in the case of an inhomogeneous earth was formerly a matter of conjecture. In particular, the result justifies (to this order of approximation) the application to each radial separately of the formulae appropriate to normal incidence on the coastline and so serves to provide a theoretical basis for the use of Clemmow's solution for a line source in a study of propagation at oblique incidence. The analytical difficulties associated with the problem of three-dimensional propagation can therefore be avoided.

As regards the present work, the above correspondence is of value in that it enables us to take over immediately the simplified

modes of solution given in Section 4.3. Thus, for a receiver on the surface of the land,

$$E_y = (1 + \rho)E_y^i \quad (16)$$

whilst for a receiver on the surface of the sea,

$$E_y = 2 \left[1 - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} (1 - \rho) F \left\{ \sqrt{\gamma(R'_1 - R)} \right\} \right] E_y^i \quad (17)$$

and this last equation may be simplified still further to read

$$E_y = \frac{2y_0/r'_0}{y_0/r'_0 + \eta} \left[1 + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma R r'_0}{r'_0} \right)} \right] E_y^i \quad (18)$$

[see eqn. (11)], providing the receiver is not too far from the coast.

As already mentioned, the problem of a vertical electric dipole is one which has been discussed by Feinberg,⁷ and eqn. (18) may be compared with his solution. For the case in which the transmitter is at a large numerical distance from the coastline, Feinberg's expression* for the ground-to-ground field over the sea is in complete agreement with eqn. (18) when the height-gain factor for an elevated transmitter is approximated as in Section 4.3.

(6) A TRANSMITTER OVER THE SEA

The method described in Section 3 does not permit a full discussion of the propagation problem in terms of the given model. A vital feature of the method is the imaging process, which requires that the transmitter be situated far over the land, whereas the various cases which can exist are

- (a) The transmitter far over land; the receiver anywhere over sea.
- (b) The receiver far over land; the transmitter anywhere over sea.
- (c) The transmitter far over sea; the receiver anywhere over land.
- (d) The receiver far over sea; the transmitter anywhere over land.

These four cases cover all possible positions of transmitter and receiver (of large distance apart and on opposite sides of the discontinuity). Of these only (a) has been treated explicitly, although (b) may be regarded as solved since its solution can be deduced from that of (a) by the reciprocity condition. Similarly, the solution for (d) can be deduced from that of (c), and attention will therefore be confined to this last case.

If (a) and (c) are compared, it is seen that they correspond to one another with land and sea interchanged, and this suggests the introduction of a model, complementary to that of Section 2, in which an imperfectly conducting sheet (representing the land) lies in the air-sea interface. Such a model gives an alternative, but equally valid, picture of the idealized coastal region and enables case (c) to be studied using a technique analogous to the above in which the presence of the sea (of reflection coefficient unity) is accounted for by the insertion of an identical image field.

The basic diffraction problem now involves a metallic sheet, and only for an incident plane wave is the solution known.¹⁸ Other types of incident field can be expressed in terms of plane waves, and whilst the main interest lies in the solution for a point source, it is sufficient to consider the two-dimensional problem of a line source in view of the correspondence referred to in Section 5.

The analysis is similar to that in Section 4 and the details will be omitted. If all slowly varying factors are removed from the

integrands at their predominant values, the total field takes the form

$$H_z = \left[1 + (\psi, 1)_{\theta_0} \varepsilon^{-j\gamma(S-R)} + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \Gamma(\theta, \theta_0) \left\{ F[\gamma(R_1 - R)]^{1/2} \pm \varepsilon^{-j\gamma(S-R)} F[\gamma(R_1 - S)]^{1/2} \right\} \right] \frac{\varepsilon^{-j\gamma R}}{\sqrt{(\gamma R)}} \quad (19)$$

with the upper or lower sign according as $\theta < \pi - \theta_0$ or $\theta > \pi - \theta_0$ respectively. The factor $\Gamma(\theta, \theta_0)$ contains the 'split' functions¹⁸ appearing in the solution for diffraction by a metallic sheet, and some simplification may be achieved by assigning to $\Gamma(\theta, \theta_0)$ the value $[1 - \rho(\sin \theta_0)]$ which it assumes on the shadow edge $\theta = \pi - \theta_0$. Comparison with the expression for H_z obtained using the previous model [eqn. (8)] then shows a modification to the geometrical-optics term and a change in sign of the diffraction term. Moreover, the conditions upon the validity of eqns. (8) and (19) are similar, and though the straightforward treatment of Γ as a polar diagram demands a slightly greater transmitter elevation, it is clear from the analysis that a more refined evaluation of the integral would enable this restriction to be relaxed without materially affecting the answer.

As in Section 4.3, a further reduction of the solution is provided by the case of a receiver at ground level. If the receiver is on the surface of the sea,

$$H_z = 2 \frac{\varepsilon^{-j\gamma R}}{\sqrt{(\gamma R)}} \quad (20)$$

a result which is consistent with the perfectly conducting nature of the sea. If the receiver is on the surface of the land, the approximation to $\Gamma(\theta, \theta_0)$ is only possible at points very near to the coast, and here the analysis of Section 4.3 can be employed to give

$$H_z = 2 \left[1 - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma r r_0}{R} \right)} \right] \frac{\varepsilon^{-j\gamma R}}{\sqrt{(\gamma R)}} \quad (21)$$

Since the height-gain factor for the transmitter is unity, eqn. (21) also represents the ground-to-ground field, and as such may be compared with that obtained by Clemmow.¹⁰ Under the conditions

$$|\eta \sqrt{(\gamma r/2)}| \ll 1 \ll |\eta \sqrt{(\gamma R/2)}|$$

Clemmow's result reduces to a form differing from eqn. (21) merely in having r_0/R replaced by unity, a justifiable approximation if r is small. This provides satisfactory confirmation of the present solution for points near to the coast, and since the case when r is not small is effectively included in case (b), it can be claimed that the analysis of the propagation problem is now complete.

(7) THE NATURE OF THE FIELD

The behaviour of the field in a region which includes the coastline can best be studied with the aid of a few numerical examples. For this purpose the transmitter is taken to be a point source over the land and at a distance of 300 wavelengths from the boundary, whilst the land itself is assumed to be a pure dielectric with $\sin \alpha_B = 1/3$.

(7.1) Field-Strength Recovery Effect

The existence of the field-strength recovery effect was first predicted by Millington^{2,19} on the basis of his technique for treating ground-wave propagation over a land-sea boundary, and striking confirmation over a wide range of frequencies was provided by a series of experiments conducted shortly after-

* Actually Feinberg treats only the problem of a transmitter on the surface of the sea, and the solution for a transmitter on the land must be deduced using the reciprocity condition.

wards.^{19, 20, 21, 22} In fact, the recovery always accompanies propagation from a medium of low to a medium of high complex refractive index (there is a corresponding *decrease* of field strength in the reverse direction) and owes its existence to a vertical redistribution of energy near the coastline; such a redistribution is inevitable in view of the different variations with height of the field on opposite sides of the boundary.

The recovery is illustrated in Fig. 3, in which the field strength is plotted for an elevated transmitter and a ground-level receiver.

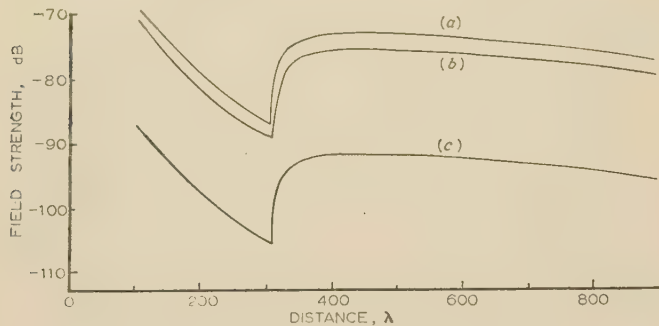


Fig. 3.—The field strength at a ground-level receiver.

- (a) Transmitter height = 4λ .
(b) Transmitter height = 3λ .
(c) Transmitter at ground level.

The upper two curves have been calculated using eqns. (16) and (17) and are for transmitter heights of 4λ and 3λ , these heights having been chosen in order to satisfy the inequalities (9). It is seen that the curves are parallel to one another except for a distance of a few wavelengths on the far side of the coast, and this gives confidence in the application of a height-gain factor for the determination of the ground-to-ground field. Accordingly, the third curve is for a ground-level transmitter, deduced from that for $y_0 = 3\lambda$ using the theoretical height-gain factor $(1 + j2\pi)^{-1}$; the agreement with the field strength obtained from Clemmow's exact solution¹⁰ is so close as to make graphical comparison impossible.

At points for which $d < 300\lambda$ the curves are merely those appropriate to a land of infinite extent, which implies that (to the present approximation) the field is undisturbed before the boundary is reached. Whilst this is also a feature of most other theoretical solutions, it is of interest to note a claim by Worledge²³ to have obtained experimental results suggesting the reflection of radio waves from the landward side of a neighbouring coastline. Since no description of the coast is given, the effect may have been due to change of relief, and, indeed, Monteath²⁴ has shown that the apparent reflection coefficient of a discontinuity in dielectric constant is so small (e.g. $|\rho| = 0.007$ for $\eta_1/\eta_2 = 0.2$) as to be negligible.

Just beyond the coastline the solutions reveal a marked increase of field strength, the initial rate of recovery being extremely rapid. Thus for a transmitter and receiver at ground level, the intensity at $r' = \lambda$ is 3.4 dB above its value at the boundary, rising to a local maximum of 13.7 dB at $r' = 130\lambda$. Sufficiently far over the sea the attenuation is characteristic of the sea itself, confirming the obvious supposition of Eckersley²⁵ and Millington.² This can be seen from Fig. 3 and also from eqn. (18): when $r'/R \approx 1$,

$$E_y = 2 \frac{e^{j\pi/4}}{\sqrt{\pi}} y_0 \sqrt{\left(\frac{2\gamma}{r'_0}\right) \frac{e^{-j\gamma R}}{\gamma R}} \quad \dots \quad (22)$$

and the field is equivalent to that of a transmitter in the presence of an infinite perfectly conducting sheet. The power and phase are modified in accordance with eqn. (22).

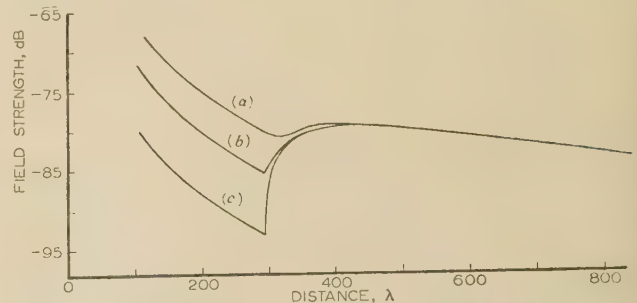


Fig. 4.—The field strength at an elevated receiver for transmitter height 2λ .

- (a) $y = 2\lambda$, (b) $y = \lambda$, (c) $y = 0$.

Corresponding curves for the receiver heights, $y = 2\lambda$, $y = \lambda$, and $y = 0$, are given in Fig. 4. The transmitter height is now 2λ and it will be observed that there is again no disturbance of the field before the coastline. As soon as the boundary is crossed the curves for an elevated receiver change their character in such a way as to join up with that for a receiver at ground level, the 'transition distance' depending on the value of y . Since the theoretical height-gain factor for the sea is unity for all receiver heights, these results once more confirm the application of height-gain everywhere except within the immediate vicinity of the coast.

(7.2) The Change in Phase

The recovery in field strength is accompanied by a notable change of phase, as shown in Fig. 5. The phase, relative to that

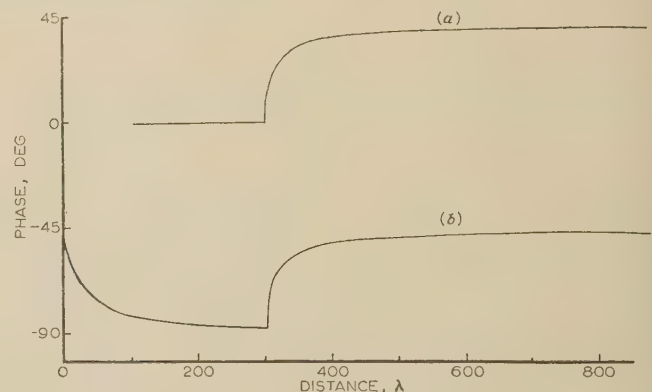


Fig. 5.—The phase relative to that in free space.

The receiver is at ground level.

- (a) Transmitter height = 4λ (present theory).
(b) Transmitter at ground level (Clemmow's theory).

of a transmitter in free space, is here plotted against distance, the lower curve, (b), having been calculated using Clemmow's solution for the ground-to-ground field. Except at points close to the transmitter, the phase over the land is roughly constant and is asymptotic to $-\pi/2$ with increasing distance from the transmitter. The phase velocity is therefore negligibly less than that of free-space propagation. The crossing of the boundary is followed by a rapid increase of phase, leading to a velocity exceeding that of free space, and it is this effect which gives rise within a few wavelengths of the boundary, to the phenomenon of coastal refraction. Outside this region the curve flattens out and its asymptotic value, $-\pi/4$, lies half-way between those appropriate to a perfectly conducting earth and a pure dielectric one.

Curve (a) of Fig. 5 has been obtained using the present theory, and is for an elevated transmitter ($y_0 = 4\lambda$) and a ground-level receiver. Over the land the phase relative to that of a transmitter in free space is exactly zero, although it should be borne in mind that the solution is valid only for the radiation field of the transmitter. The curve thus differs by almost $-\pi/2$ from (b) and this is effectively the phase of the theoretical height-gain factor $(1 + j8\pi/3)^{-1}$ which must be applied to the transmitter in order to determine the ground-to-ground field. At points just beyond the coast the change of phase is identical in both curves, whilst far out over the sea the phase is asymptotic to $41^\circ 42'$, exceeding that of Clemmow's solution by $86^\circ 42'$.

For other transmitter and receiver heights which are in accordance with the inequality (9d), the phase curves derived from eqn. (15) differ negligibly from (a). Thus, for $y_0 = 3\lambda$, $y = 0$, the curve is identical with the above over land and within 1° of it over the sea; and change of receiver height has similarly little effect. Phase plots additional to (a) are therefore unnecessary.

(7.3) Height-Gain Considerations

From the analysis for propagation over a homogeneous earth (e.g. Reference 26), it is possible to derive an expression for the ratio of the field at a height h to that on the surface of the earth. Such an expression is known as a height-gain factor and has been shown by Eckersley and Millington^{27, 28} to be of the same form, namely $(1 + j\gamma h \sin \alpha_B)$, on both flat- and curved-earth theories.

Although normally used only in regard to field strength, height-gain factors are also applicable to phase. It follows that the field for an elevated transmitter and receiver may be written as

$$E_y = (1 + j\gamma y_0 \sin \alpha_B)(1 + j\gamma y \sin \alpha_B)E''_y \quad (23)$$

where E''_y is the ground-to-ground field, and whilst it is not easy to judge precisely up to what heights this is valid, a criterion which has been suggested is

$$\gamma^2(y_0^2 + y^2) \ll \gamma d \quad (24)$$

It is natural to suppose that under certain conditions these factors may be used in the mixed-path problem where, in the examples taken, $\sin \alpha_B = 1/3$ for the land and zero for the sea. Their application to phase may be studied by reference to Fig. 5. The theoretical factor required to reduce the transmitter height of curve (a) to zero [curve (b)] is here $(1 + j8\pi/3)^{-1}$, the phase of which is $83^\circ 11'$, but this does not completely align the two curves, particularly over land. If the inequality (24) is operative, however (and this is a reasonable assumption when the transmitter and receiver are over the same medium), the values $y_0 = 4\lambda$, $y = 0$ restrict the application of a height-gain factor to those positions of the receiver for which $d > 275\lambda$ (say), and here (a) may be brought into coincidence with (b) using a factor whose argument is $-\pi/2$. Moreover, all phase curves deduced from eqn. (15) are essentially the same as (a) for transmitter heights satisfying the conditions (9), and the argument must therefore be independent of y_0 . This is obtained if $(1 + j\gamma y_0 \sin \alpha_B)$ is replaced by $j\gamma y_0 \sin \alpha_B$, an approximation consistent with (9d), and if the height-gain factor for the transmitter is taken to be of this form, the phase curve for the ground-to-ground field differs by at most 3° from that of Clemmow.

Consider now the application of the height-gain factor to the receiver. It was noted in Section 7.2 that the phase is substantially unaffected by change of receiver height, so that a height-gain factor which is purely real is required to align the phase curve for (say) $y = 2\lambda$ with that for $y = 0$. Whilst this corresponds to the theoretical factor for positions of the receiver over the sea, it certainly does not when the receiver lies over the land. If the appropriate factor is applied according to the position of

the receiver, it is found that the discontinuity in the argument of the factor leads to a discontinuity in the resulting phase curve, and under these circumstances it is clear that height-gain considerations are inapplicable.

The reason for this distinction between the transmitter and the receiver is easily seen. In the examples given, the receiver is allowed to cross the coastline, whereas the transmitter remains constantly over the same medium, and if the use of a height-gain factor is confined to points which are far from the boundary, sensible results are obtained even for the receiver.

In connection with field strength the application of height-gain to the transmitter is straightforward, a comparison of curves (a) and (b) of Fig. 3 being sufficient to verify this; application to the receiver, however, calls for more careful attention. In Fig. 6

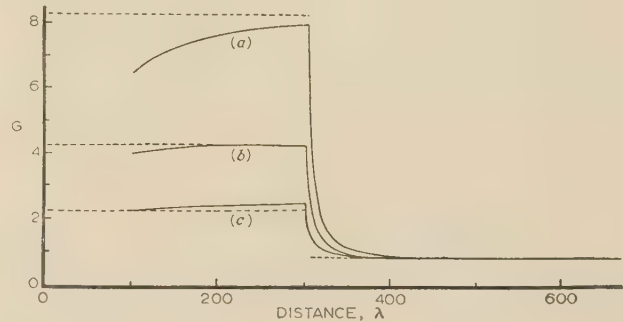


Fig. 6.—The modulus, G , of the calculated height-gain factor required to reduce the receiver height from y to zero.

The transmitter height has a constant value of 4λ .

(a) $y = 4\lambda$; (b) $y = 2\lambda$; (c) $y = 0$.

Dotted lines represent moduli of corresponding theoretical height-gain factors for homogeneous earths composed of each medium separately.

the ratio, G , of the field strength at a height y to that on the surface of the earth is plotted against distance, the transmitter height being kept constant at 4λ . The curves have been obtained from eqn. (15), and, in effect, they represent the moduli of the required height-gain factors, as opposed to the theoretical factors whose moduli are also shown.

For $y = \lambda$ the calculated factor is in close agreement with the theoretical one, the only important difference being over the first 25 wavelengths of the sea path. Similarly for $y = 2\lambda$, but in this case there is a tendency for the theoretical factor to fail at distances from the transmitter of less than 150 wavelengths. When $y = 4\lambda$ this failure becomes more pronounced, and over the entire land path the theoretical factor departs noticeably from the calculated one, the discrepancy varying from 2 at $d = 100\lambda$ to $\frac{1}{2}$ at $d = 300\lambda$. Over the sea, however, the theoretical factor is, in the main, sufficient except for the first 80 wavelengths of the sea path.

In the cases quoted the failure of the theoretical factor over the land may be attributed to the violation of condition (24). This is particularly so for $y = 4\lambda$, when the transmitter and receiver heights exceed those for which height-gain is valid, and if such values of y_0 and y are excluded, the theoretical height-gain factor is applicable to field strength everywhere except within a strip bordering on the coast.

(8) CONCLUDING REMARKS

The attention which has been directed at the mixed-path problem in recent years has led to the rigorous derivation of solutions in substantial agreement with one another. The complications of analysis, however, are in direct contrast to the relatively simple forms to which the solutions reduce in many cases of practical interest, and it was this fact which suggested

the possibility of a more elementary treatment. Since the main feature of the problem is the existence of a diffracting coastline, a method was developed which achieves simplicity by calling upon the known results for diffraction at a straight edge, thereby eliminating the more difficult analysis. The solutions obtained in this manner reveal all the important features of mixed-path propagation and, moreover, are in a form suitable for numerical application.

Although the method has been described in terms of an idealized problem in which one medium is perfectly conducting, it is clear that this restriction can be removed by using the corresponding results for diffraction by a sheet of finite conductivity. The general problem is therefore susceptible to the new technique.

The remaining assumptions are common to all previous treatments and are that the earth is smooth and flat with only a single discontinuity in dielectric constant. Whilst these conditions are essential for the success of the method, other discontinuities can be admitted providing their separation is large enough to allow the propagation to become characteristic of each individual medium; in this event the boundaries may be considered one by one.

(9) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. P. C. Clemmow of Cambridge University for much assistance and advice and to acknowledge the receipt of a grant from the Department of Scientific and Industrial Research for the period during which this work was carried out.

(10) REFERENCES

- (1) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *Journal of Physics USSR*, 1944, **8**, p. 317.
- (2) MILLINGTON, G.: 'Ground-Wave Propagation over an Inhomogeneous Smooth Earth', *Proceedings I.E.E.*, Paper No. 794 R, January, 1949 (**96**, Part III, p. 53).
- (3) ECKERSLEY, T. L.: Atti del Congresso Internazionale per il Cinquantenario della Scoperta Marconiana della Radio, Rome, 1947, p. 78.
- (4) PRESSEY, B. G., ASHWELL, G. E., and FOWLER, C. S.: 'The Measurement of the Phase Velocity of Ground-Wave Propagation at Low Frequencies over a Land Path', *Proceedings I.E.E.*, Paper No. 1438 R, March, 1953 (**100**, Part III, p. 73).
- (5) GRÜNBERG, G.: 'Theory of the Coastal Refraction of Electromagnetic Waves', *Journal of Physics USSR*, 1942, **6**, p. 185.
- (6) GRÜNBERG, G.: 'Suggestions for a Theory of Coastal Refraction', *Physical Review*, 1943, **63**, p. 185.
- (7) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *Journal of Physics USSR*, 1946, **10**, p. 410.
- (8) FURUTSU, K.: 'Propagation of Electromagnetic Waves over a Flat Earth across a Boundary separating Different Media and Coastal Refraction', *Journal of the Radio Research Laboratories*, Tokyo, Japan, 1955, **2**, p. 1.
- (9) AONO, Y., and MURAMATSU, K.: 'Measurement of Field Intensity of Ground Wave over Mixed Paths', *ibid.*, p. 51.
- (10) CLEMMOW, P. C.: 'Radio Propagation over a Flat Earth across a Boundary separating Two Different Media', *Philosophical Transactions of the Royal Society, A*, 1953, **246**, p. 1.
- (11) HÖLLER, P.: 'Zur Ausbreitung elektromagnetischer Wellen von Land nach See und umgekehrt (Teil 1)', *Zeitschrift für angewandte Physik*, Berlin, 1951, **3**, p. 424.
- (12) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *Journal of Physics USSR*, 1945, **9**, p. 1.
- (13) CLEMMOW, P. C.: 'A Method for the Exact Solution of a Class of Two-Dimensional Diffraction Problems', *Proceedings of the Royal Society, A*, 1951, **205**, p. 286.
- (14) RANKIN, R. A.: 'The Mathematical Theory of the Motion of Rotated and Unrotated Rockets', *Philosophical Transactions of the Royal Society, A*, 1949, **241**, p. 457.
- (15) CLEMMOW, P. C.: 'A Note on the Diffraction of a Cylindrical Wave by a Perfectly Conducting Half-Plane', *Quarterly Journal of Mechanics and Applied Mathematics*, 1950, **3**, p. 377.
- (16) SENIOR, T. B. A.: 'The Diffraction of a Dipole Field by a Perfectly Conducting Half-Plane', *ibid.*, 1953, **6**, p. 101.
- (17) BOOKER, H. G., and CLEMMOW, P. C.: 'A Relation between the Sommerfeld Theory of Radio Propagation over a Flat Earth and the Theory of Diffraction at a Straight Edge', *Proceedings I.E.E.*, Paper No. 873 R, January, 1950 (**97**, Part III, p. 18).
- (18) SENIOR, T. B. A.: 'Diffraction by a Semi-Infinite Metallic Sheet', *Proceedings of the Royal Society, A*, 1952, **213**, p. 436.
- (19) MILLINGTON, G.: 'Ground-Wave Propagation across a Land/Sea Boundary', *Nature*, 1949, **163**, p. 128.
- (20) MILLINGTON, G.: 'Ground-Wave Propagation across a Land/Sea Boundary. 100m Waves', *ibid.*, 1949, **164**, p. 114.
- (21) ELSON, N.: 'Ground-Wave Propagation across a Land/Sea Boundary. 300m Waves', *ibid.*, p. 114.
- (22) MILLINGTON, G., and ISTEAD, G. A.: 'Ground-Wave Propagation over an Inhomogeneous Smooth Earth. Part 2. Experimental Evidence and Practical Implications', *Proceedings I.E.E.*, Paper No. 909 R, February, 1950 (**97**, Part III, p. 209).
- (23) WORLLEDGE, J. P. G.: 'Deviation of Wireless Waves at a Coastal Boundary', *Nature*, 1928, **121**, p. 35.
- (24) MONTEATH, G. D.: 'Application of the Compensation Theorem to Certain Radiation and Propagation Problems', *Proceedings I.E.E.*, Monograph No. 3, June, 1951 (**98**, Part IV, p. 23).
- (25) ECKERSLEY, P. P.: 'The Calculation of the Service Area of Broadcast Stations', *Proceedings of the Institute of Radio Engineers*, 1930, **18**, p. 1160.
- (26) NORTON, K. A.: 'The Calculation of Ground-Wave Field Intensity over a Finitely Conducting Spherical Earth', *ibid.*, 1941, **29**, p. 623.
- (27) ECKERSLEY, T. L., and MILLINGTON, G.: 'Application of the Phase Integral Method to the Analysis of the Diffraction and Refraction of Radio Waves round the Earth', *Philosophical Transactions of the Royal Society, A*, 1938, **237**, p. 273.
- (28) ECKERSLEY, T. L., and MILLINGTON, G.: 'The Experimental Verification of the Diffraction Analysis of the Relation between Height and Gain for Radio Waves of Medium Lengths', *Proceedings of the Physical Society*, 1939, **51**, p. 805.

(11) APPENDIX

(11.1) The Diffraction of a Three-dimensional Plane Wave by a Half-Plane

This is merely an extension of the classical Sommerfeld diffraction problem and its solution can be deduced therefrom. It is clear that any two-dimensional solution of the wave equation gives rise to a quasi three-dimensional solution or

replacing γ by $\gamma \cos \beta$ and multiplying by $\exp(-j\gamma z \sin \beta)$. If V is such a three-dimensional solution, an electromagnetic field may be derived from it by taking V as the z component of an electric or magnetic Hertz vector whose x and y components are zero. This serves to produce two fundamental types of field:

E-polarized in which $H_z = 0$,

$$\begin{aligned} E^{(1)} &= \left(\frac{j \sin \beta}{\gamma} \frac{\partial V}{\partial x}, \frac{j \sin \beta}{\gamma} \frac{\partial V}{\partial y}, -\cos^2 \beta V \right), \\ H^{(1)} &= \frac{jY}{\gamma} \left(-\frac{\partial V}{\partial y}, \frac{\partial V}{\partial x}, 0 \right). \end{aligned} \quad (25)$$

H-polarized in which $E_z = 0$,

$$\begin{aligned} E^{(2)} &= \frac{jZ}{\gamma} \left(-\frac{\partial V}{\partial y}, \frac{\partial V}{\partial x}, 0 \right), \\ H^{(2)} &= \left(-\frac{j \sin \beta}{\gamma} \frac{\partial V}{\partial x}, -\frac{j \sin \beta}{\gamma} \frac{\partial V}{\partial y}, \cos^2 \beta V \right). \end{aligned} \quad (26)$$

Consider that solution of the wave equation corresponding in the above manner to $\exp[j\gamma(x \cos \alpha + y \sin \alpha)]$. It is

$$w = \exp[j\gamma(x \cos \alpha \cos \beta + y \sin \alpha \cos \beta - z \sin \beta)]$$

in terms of which

$$\begin{aligned} E^{(1)} &= (-\cos \alpha \sin \beta \cos \beta, -\sin \alpha \sin \beta \cos \beta, -\cos^2 \beta)w \\ H^{(1)} &= (Y \sin \alpha \cos \beta, -Y \cos \alpha \cos \beta, 0)w \end{aligned} \quad (27)$$

$$\begin{aligned} E^{(2)} &= (Z \sin \alpha \cos \beta, -Z \cos \alpha \cos \beta, 0)w \\ H^{(2)} &= (\cos \alpha \sin \beta \cos \beta, \sin \alpha \sin \beta \cos \beta, \cos^2 \beta)w \end{aligned} \quad (28)$$

When $\beta = 0$ these reduce respectively to *E*- and *H*-polarized fields in two dimensions. This suggests that the solutions of the diffraction problem for the incident fields [eqns. (27) and (28)] may be deduced from those in the case $\beta = 0$ —a procedure which is certainly valid if the diffracting structure is two-dimensional. Thus, take the two-dimensional solution E_z , replace γ by $\gamma \cos \beta$ and multiply by $\exp(-j\gamma z \sin \beta)$. When substituted into eqn. (25) in place of V , this forms the solution of the diffraction problem for the incident field [eqn. (27)]. And similarly for *H*-polarization.

Let us now specify the incident field to be one whose magnetic vector has no component normal to the sheet. If the fields of eqns. (27) and (28) are combined using constants p and q , the latter must be chosen such that

$$H_y = pH_y^{(1)} + qH_y^{(2)} = 0$$

a condition which is satisfied by $p = \sin \alpha \sin \beta$ and $q = Y \cos \alpha$. The resulting field is then

$$\begin{aligned} E &= [\sin \alpha \cos \alpha \cos^3 \beta, -\cos \beta(1 - \sin^2 \alpha \cos^2 \beta), \\ &\quad -\sin \alpha \sin \beta \cos^2 \beta]w \\ H &= [Y \sin \beta \cos \beta, 0, Y \cos \alpha \cos^2 \beta]w \end{aligned} \quad (29)$$

If this is incident upon the half-plane, the total field may be found by inserting into the appropriate combination of eqns. (25) and (26) the solutions of the Sommerfeld problem for two-dimensional *E*- and *H*-polarized fields, providing these are modified in the manner described above. When the differentiations are carried out

$$\begin{aligned} E_x &= \sin \alpha \cos \alpha \cos^2 \beta(F_1 - F_2) \\ &\quad + [1 - (1 - \cos \alpha) \cos^2 \beta] \tan \frac{1}{2} \theta G \\ E_y &= -(1 - \sin^2 \alpha \cos^2 \beta)(F_1 + F_2) \\ &\quad - [1 - (1 - \cos \alpha) \cos^2 \beta]G \\ E_z &= -\sin \alpha \sin \beta \cos \beta(F_1 - F_2) \\ H_x &= Y \sin \beta(F_1 + F_2) + Y \sin \beta G \\ H_y &= Y \tan \beta \tan \frac{1}{2} \theta G \\ H_z &= Y \cos \alpha \cos \beta(F_1 + F_2) \end{aligned} \quad (30)$$

where

$$G = \cos \frac{1}{2} \alpha \cos \frac{1}{2} \theta \sqrt{\left(\frac{2 \cos \beta}{\pi \gamma r} \right)} \exp[-j\pi/4 - j\gamma(r \cos \beta + z \sin \beta)]$$

and

$$F_1 \equiv F_1(\alpha) = \frac{1}{\sqrt{\pi}} \cos \beta \exp(-j\pi/4)w \times F[-\sqrt{(2\gamma r \cos \beta)} \cos \frac{1}{2}(\theta - \alpha)]$$

with $F_2 = F_1(-\alpha)$. These expressions represent the required field.

CHARACTERISTICS OF THE TRIGATRON SPARK-GAP

By A. M. SLETTEN, Sivilingeniør, M.N.I.F., and T. J. LEWIS, M.Sc., Ph.D., Associate Member.

(The paper was first received 27th January, and in revised form 26th May, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

The behaviour of a trigatron three-electrode spark-gap in air has been investigated and its characteristics obtained with particular reference to its use as a controlled high-voltage switch. It is found that the voltage range over which it may be triggered satisfactorily depends, not only on the polarities of the main gap and triggering voltages, but also on the energy of the discharge. The breakdown time-lag is also determined by these same voltage polarities and also by the time-constant of the trigger discharge circuit.

From these characteristics and certain other relevant observations, a theory of the breakdown process in such a spark-gap is suggested, involving the propagation from the trigger of a low-density easily-ionized region.

Finally, a brief investigation of the successful use of a trigatron in a diverter circuit capable of diverting the discharge energy in a spark-gap subjected to direct and impulse voltages of 300kV and in a circuit providing accurate 'chopping' of impulse voltage waves, is reported.

(1) INTRODUCTION

There is a frequent need for a high-voltage switch capable of rapid and controlled operation over a considerable voltage range. This need arises, for instance, in the operation of impulse generators, in the production of 'chopped' impulse voltages and also in the diversion of the discharge energy when electrical breakdown occurs in equipment under test. In the latter case, the energy dissipated may frequently be destructive, preventing further tests on the particular test object. For instance, in experiments to determine the electric strength of an insulating oil, each breakdown causes a deterioration of the insulating properties and limits the number of measurements possible on a sample.

Below about 10kV, the various switching operations outlined above can be performed by thyatrons. Using such devices, the discharge energy of a breakdown, for example, may be diverted from the test object in times of less than 1 microsec.¹ Impulse-generator control also presents no difficulties when thyatrons are used. Above this voltage, it is usually necessary to resort to a spark-gap in air as the switching device, and to provide a third electrode as the control or trigger. One such three-electrode spark-gap is the trigatron,² consisting of two hemispherical main electrodes, together with a trigger electrode of small diameter mounted centrally in a hole in one of the main electrodes, and insulated from it (see Fig. 1). A comparatively small voltage applied to the trigger electrode can cause the main gap to break down at voltages below the normal sparking voltage. The device is thus a controlled switch. The development of trigatrons for use as modulators in recurrent-pulse equipment, in which the requirements are those of high repetition rate and freedom from jitter, has been well described by Craggs, Haine and Meek,³ and also by Wilkinson,⁴ whilst Husbands and Higham⁵ and Hardy and Broadbent¹² have described circuits employing the trigatron to control impulse generators.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Sletten is with Standard Telefon og Kabelfabrik, Norway, and was formerly in the Electrical Engineering Department, Queen Mary College, London University. Dr. Lewis is in the Electrical Engineering Department, Queen Mary College, London University.

The authors sought to use the trigatron as an energy diverter in circuits operating at voltages up to about 300kV, for which purpose it became necessary to know the operating range and the inherent time lag of breakdown of trigatrons for which the normal untriggered static breakdown voltage was 100kV. Specific information on these matters, which is not provided in the References mentioned, has been obtained, and also permits a theory of breakdown in such a device to be attempted. Although the scope of the present investigation is limited, so that the theory must be tentative, it may be of special interest since breakdown is caused in a spark-gap, which, initially at least, is considerably under-stressed. Under such conditions, the usual processes of ionization and avalanche formation are not occurring naturally before the beginning of breakdown as they would be for the static breakdown of the corresponding two-electrode gap.

Using the trigatron characteristics obtained, tests were made on a complete diverter incorporating it, in circuits up to 300kV. The results indicate that an efficient diverter or 'chopping' device can be constructed with the trigatron.

(2) THE TRIGATRON

The trigatron tested (see Fig. 1) consisted of two 15 cm-diameter aluminium hemispheres mounted horizontally with an adjustable gap. A $\frac{1}{8}$ -in-diameter steel rod formed the trigger electrode, and this was insulated from the main electrode by a close-fitting glass tube. The trigger electrode and the glass tube could be adjusted to provide different settings, but, in all the tests to be described, the glass tube was set flush with the main electrode, as in Fig. 1. Unless otherwise stated, the trigger electrode was

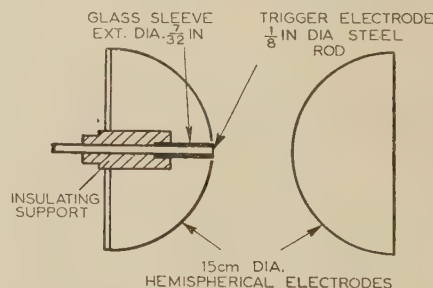


Fig. 1.—Trigatron spark gap.

also mounted flush with the main electrode, giving a trigger discharge gap of approximately 1.2 mm. Unlike previous trigatron designs, no outer conducting surface was applied to the glass tube; subsequent results showed that such a surface would not affect the breakdown significantly.

The trigatron was mounted in air at atmospheric pressure, and no attempt was made to control the atmospheric conditions, so that some of the variations in the results to be described might be due to this cause. However, more precise control did not seem justified for the present investigation.

The full diverter, tests on which are reported in Section 6, consisted of three pairs of hemispherical electrodes connected in series and mounted one above the other (see Fig. 11). The lower

pair was modified to form the trigatron as above. The three gaps are arranged to be in line, so that the upper pairs are irradiated from the trigatron discharge, producing a short statistical time lag. For direct-voltage operation each spark-gap had a 170-megohm resistor connected across it to ensure a uniform voltage distribution. Capacitors of $250\ \mu\text{F}$ capacitance were connected across the resistors when operating with impulse voltages.

(3) TEST EQUIPMENT

The trigatron characteristics have been determined with direct voltage having a maximum value of 100 kV across the main gap throughout. The circuit arrangements are shown in Fig. 2, in

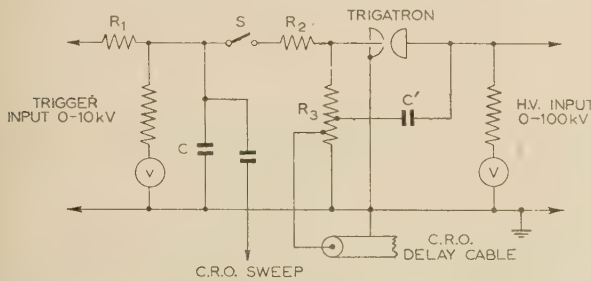


Fig. 2.—Circuit arrangements for recording trigatron characteristics.

which a high-voltage resistor and microammeter in series determined the main gap voltage V . The trigger-gap discharge was initiated by a voltage impulse from a separate source consisting of a capacitor C charged through a high resistance R_1 from a d.c. supply having a maximum output of 10 kV. The voltage v on C could be applied to the trigger electrode through the resistor R_2 by closing a remotely controlled switch S , the trigger electrode itself being earthed through a 1-megohm resistor R_3 . Since the drilled electrode of the trigatron was permanently earthed, it will be convenient to refer the voltages V and v to this common point. According to the polarities of these voltages there will be four possible test combinations.

The trigger pulse was recorded on an oscilloscope by means of a low-voltage tapping on the resistor R_3 . The sudden collapse in the voltage V , which marks the breakdown of the main gap, was also recorded through the same delay circuit being fed through the small capacitor C' from the high-voltage electrode. The collapse of voltage on C on closing the switch S was utilized to initiate the time sweep of the oscilloscope. A typical record showing the trigger pulse and the subsequent sudden potential change corresponding to breakdown of the main gap is shown in

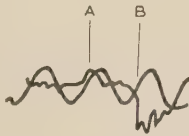


Fig. 3.—Typical time-lag oscillogram.

The interval AB marks the time lag between the trigger spark and the breakdown of the main gap.

Gap length = 3.5 cm.
 $C = 0.04\ \mu\text{F}$.
 $R_2 = 0$.
 $V = -65\ \text{kV}$.
 $v = -10\ \text{kV}$.
 Timing oscillation = 2 Mc/s.

Fig. 3. The breakdown time-lag was measured as the interval AB between the two peaks of v and V . Any error involved in this arbitrary choice will not materially alter the general deductions from these measurements. In the same way, some unwanted oscillations also present on the record might cause small errors.

These oscillations arose in the recording circuits, but it was not deemed necessary to eliminate them completely.

(4) EXPERIMENTAL RESULTS

(4.1) Static Breakdown Voltage

The static breakdown voltage V_s of the trigatron gap, i.e. the breakdown voltage in the absence of a trigger pulse, proved to be the same as that for a normal 15 cm-diameter sphere-gap when the trigger electrode was flush with the main electrode. No change in V_s was detected when the trigger electrode was withdrawn behind the main electrode surface, but it dropped sharply when the trigger projected. There was also no change in V_s when R_3 was varied, provided that its value was several thousand ohms or less.

(4.2) Lower Triggering Limit

The usefulness of the trigatron as a switching device depends on the voltage range over which it may be triggered efficiently, and therefore the lower limit of operation or triggering limit V_l must be determined. When V was positive, V_l was well defined and the same for both polarities of v , but when V was negative the breakdown voltage in the neighbourhood of the triggering limit appeared to be subject to statistical fluctuations, as shown in Fig. 4. These results were obtained by observing the number of

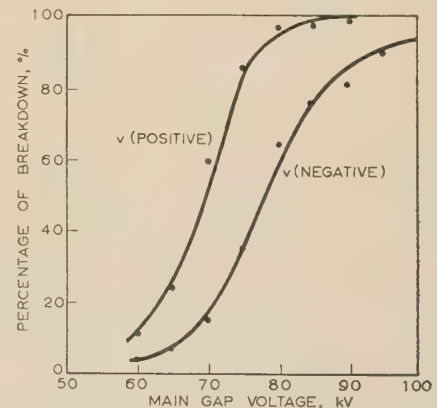


Fig. 4.—Statistical distributions of breakdowns for negative main gap.

Gap length = 7 cm.
 $C = 0.04\ \mu\text{F}$.
 $R_2 = 0$.
 $v = 10\ \text{kV}$.

breakdowns resulting from 100 trigger pulses at each main-gap voltage V . Such distributions, which appear to be normal, were observed over the whole range of gap settings up to 7 cm. When V was negative, V_l (taken to be the lowest value of V for which ten successive trigger pulses caused breakdown) was dependent on the polarity of v —being lower when v was positive. The above differences between V positive and V negative are similar to those reported by Hardy and Broadbent,¹² and are not easy to explain although the time-lag measurements to be given later indicate possible reasons.

Apart from these polarity effects, V_l was greatly influenced by conditions in the trigger circuit. For instance, if C were fixed and v increased, V_l was decreased correspondingly, whilst if v were fixed and C increased, there was again a decrease in V_l . These results, together with the fact that V was unaffected by variation of R_2 over a wide range, strongly suggested that V_l was determined by the charge Q on the trigger capacitor C prior to operation. To confirm this, C and v were varied, but their product Q remained constant and V_l was measured for positive main-gap polarity at a gap of 3.5 cm. The results given in

Table 1

VALUES OF V_I (IN KILOVOLTS) FOR VARIOUS TRIGGER CAPACITANCES C AND TRIGGER CHARGES Q ($R_2 = 0$)

C	Q = Cv(μF × kV)		
	0.2	0.4	0.8
μF	kV	kV	kV
0.02	61	58	—
0.04	61	59	—
0.08	63	59	55
0.16	—	59	54

Table 1 clearly confirm the dependence of V_I on Q , and provide an important indication of the mechanism of main-gap breakdown.

V_I is also determined by the main-gap setting. Keeping the trigger circuit constant with $C = 0.04 \mu\text{F}$, $R_2 = 0$ and $v = 10 \text{ kV}$, the complete characteristics for all polarity combinations were obtained and are given in Fig. 5.

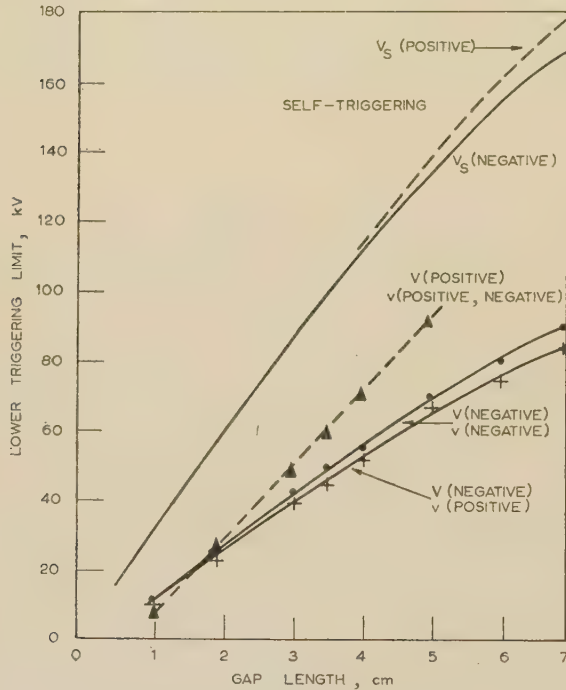


Fig. 5.—Lower triggering limit V_I .
Characteristics for all polarity combinations.
 $C = 0.04 \mu\text{F}$.
 $R_2 = 0$.
 $v = 10 \text{ kV}$ (throughout).

Certain other tests, having special significance in the discussion to follow, were made. For instance, no triggering action occurred unless a trigger discharge was actually produced. The trigger gap broke down at 2 kV , and unless v was greater than this, the main gap did not break down below V_s for any polarity combination. Thus the potential change produced by the trigger pulse is not sufficient alone to upset the gap, but the establishment of a trigger spark provides the triggering factor—a conclusion also reached by Husbands and Higham.⁵ This fact is fundamental in establishing the trigatron breakdown mechanism, especially near V_I . It should also be noted that V_I was unaffected when the trigger electrode was withdrawn 2 mm behind the face of the main electrode, provided that a trigger spark

occurred. Interesting effects were produced when a paper screen was interposed between the main electrodes. If the screen were placed close to the main electrode remote from the trigger, there was no change in V_I for any polarity combination, the screen being punctured in the process of breakdown. In any other position V_I was increased by the presence of the screen, particularly when it was close to the trigger gap shielding the main-gap volume completely. In this position the gap could not be triggered at all when V was positive, whilst when V was negative the trigger efficiency was reduced appreciably, so that $V_I \rightarrow V_s$. Lastly it was found that irradiation of the main gap by a cobalt 60γ -ray source did not change V_I , which suggests that the extra electrons introduced into the gap did not alter the breakdown process even when the gap was subjected to a voltage considerably less than V_s .

(4.3) Time Lag

The other factor determining the efficiency of the trigatron as a high-voltage switch is the operating time-lag T . Usually the breakdown time of the trigger gap will be very small and the overall time lag will be the interval AB in Fig. 3, as mentioned earlier. Measurements of T for gap lengths of 3.5 and 2 cm, for various polarity combinations, are given in Figs. 6 and 7. Constant trigger conditions ($C = 0.04 \mu\text{F}$, $R_2 = 0$ and $v = 10 \text{ kV}$) were maintained throughout the tests and the inductive loop in this circuit was also kept constant. Although considerable variation in T was found in some cases, nevertheless the general behaviour was repeatable and it is considered that the smooth curves drawn in Figs. 6 and 7 are representative. The most

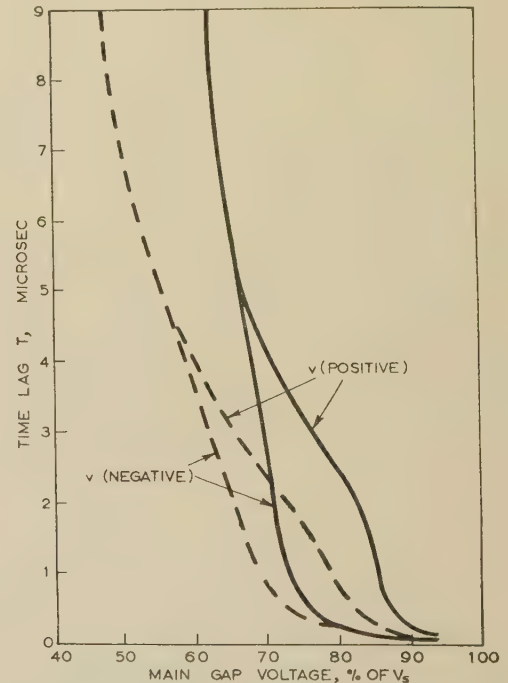


Fig. 6.—Time-lag characteristics for main gap voltage, V positive.

— 3.5 cm gap.
--- 2 cm gap.
 $C = 0.04 \mu\text{F}$.
 $R_2 = 0$.
 $v = 10 \text{ kV}$.

significant feature of the measurements with V positive (Fig. 6) was the rapid increase in T as V was decreased when v was positive. This rapid increase did not occur when v was negative until V_I was approached.¹² At this limit T increased rapidly, approaching the limit asymptotically for both polarities of v .

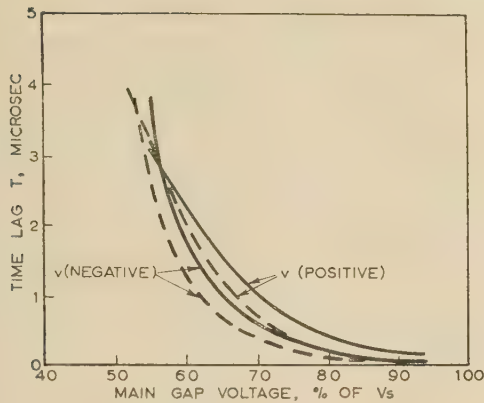


Fig. 7.—Time-lag characteristics for main gap voltage, V negative.

— 3.5 cm gap.
 --- 2 cm gap.
 $C = 0.04 \mu\text{F}$.
 $R_2 = 0$.
 $v = 10 \text{ kV}$.

Little fluctuation in T occurred for constant V , when v was positive over the whole range $V_s - V_l$, which was in contrast to the fluctuations when v was negative. In this latter case the variation in T was most evident when V was about $0.75V_s$, i.e. at the bend in the time-lag curve (Fig. 6), but it disappeared as V_l was approached, T becoming well defined in conformity with the definite nature of V_l . Of importance for the argument in the next Section is the fact that, except near V_s , T is significantly greater than the estimated time of discharge of the trigger gap, which is less than 0.25 microsec for the conditions used. Thus the influence of the trigger discharge on breakdown can be delayed considerably.

When V is negative, T is considerably less for all polarities (Fig. 7), except perhaps near V_l , in which region T increases rapidly. Two unexpected results are found; T is greater when v is positive and is only well defined when v is negative—at least when V is greater than V_l . This is in contrast to the previous results with V positive, in which T exhibited little fluctuation when v was positive. These apparently unsystematic results, which are in qualitative agreement with Hardy and Broadbent,¹² require reasonable explanation in any general theory of trigatron behaviour. Reference to Figs. 5 and 7 will show that the time-lag characteristics must cross near V_l , since V_l is lower when v is positive, and at the same time, at this limit, T should become less defined when v is negative. Unfortunately, owing to the very rapid increase in T near V_l , it is difficult to investigate the crossing point experimentally.

Since the trigger-circuit parameters C and v influenced V considerably, it is to be expected that they would influence T also. Extensive measurements revealed no influence, however, over the greater part of the range $V_s - V_l$. Only in the immediate neighbourhood of V_l could any effect be detected. For instance, at $V = 0.75V_s$, variation in C between the limits 0.02 and $0.08 \mu\text{F}$, and in v between 2 and 10 kV , with a gap of 3.5 cm , produced no change. In complete contrast, change of R_2 (Fig. 2) appreciably altered T , whereas it had already been found to cause no change in V_l . R_2 determines the time-constant of the trigger-circuit discharge, and therefore the rate of growth of the trigger spark channel. Thus a slow discharge in this circuit might give a long time-lag, at least near V_l , depending on the mechanism of main-gap breakdown. Increase in the time lag with increase in R_2 was found for all polarity combinations, the increase being negligible near V_s but growing in magnitude towards V_l . A typical result for positive values of V and v is shown in Fig. 8. Since, for efficient operation as a switch, it is



Fig. 8.—Variation of time lag with trigger-circuit resistance R_2 for positive main gap and trigger pulse.

Gap length = 3.5 cm .
 $C = 0.04 \mu\text{F}$.
 $v = 10 \text{ kV}$.

of major importance to reduce the time lag to a minimum, few observations apart from those described here have been made with resistance in the trigger circuit, and all other measurements of T reported are with $R_2 = 0$.

Corresponding to the effects on V_l , no influence on T was observed when the gap was irradiated by the cobalt source or when the trigger electrode was withdrawn behind the main electrode.

(5) THE MECHANISM OF BREAKDOWN

The process of breakdown in the trigatron at potentials below V_s appears to be complex and is not understood easily. An attempt will be made here to suggest the main factors in the discharge processes for the various polarity combinations, but no comprehensive theory seems possible without a fuller experimental investigation.

(5.1) The Trigger Spark

The most striking result is that the trigatron can be made to break down at potentials which are greater than 50% below the normal value of V_s for the corresponding two-electrode gap. Furthermore, this happens when the greater portion of the gap volume is under normal conditions of temperature and pressure and at times long after the initiating trigger discharge has ceased. Consideration of the accepted breakdown process occurring at V_s , in which primary electron avalanches occur with a collision ionization coefficient α , indicates that α would be insignificant at potentials approaching V_l if atmospheric conditions are assumed. Insufficient ionization in this way would also fail to produce any space-charge field distortion. Reluctance to discard the α process as a fundamental process in the main gap makes it necessary to seek ways in which at least local changes can cause suitable field and gas conditions for significant values of α to be attained. Since the trigger discharge is a short-duration spark in which

considerable energy dissipation occurs, it is possible that this induces suitable conditions in the main gap.

The main products of this trigger spark are as follows:

(a) A local region of hot gas of low density, corresponding to the spark channel, which would be ejected into the main gap and would continue to move after the trigger discharge had ceased. The initial growth of the spark channel can be about 10^5 cm/sec or greater, and is accompanied by a shock wave.⁶ Behind this, the low-density region would tend to propagate at lower velocities.

(b) Ionization products, electrons and positive ions. The electrons may not be so important as the relatively immobile positive ions which can produce important space charges. These charged particles may be ejected along with the hot-gas region.

(c) Photons which will irradiate the main gap. These will be very important in ensuring an initial supply of electrons, particularly at the opposing main electrode. The efficacy of spark illumination in photo-electron production is well known.⁷

(d) Hot electrode spots, especially at the trigger cathode, which will continue to emit electrons for an appreciable time after the trigger discharge has ceased.

A combination of these processes might be expected to produce significant values of α , at least in the low-density region of expelled gas, and it is then possible to give a reasonable explanation for the majority of the effects found. The essential feature that permits breakdown below V_s appears to be the injection of the hot low-density region into the gap. Such a region is likely to persist for long periods (about 10^{-4} sec), as is found from studies of dielectric recovery following arc extinction in circuit interruption.⁸ It provides a region in which ionization by electrons can occur readily and one in which positive-ion space charge builds up. It is suggested that the production of a space charge in this way, which moves out as the low-density region moves, causes the onset of a main-gap breakdown, which itself proceeds by a more normal breakdown process. The simultaneous supply of photo-electrons aids this process considerably—in fact, at potentials just below V_s the photo-electrons alone would cause breakdown. Triggering of spark-gaps by photons has been reported at potentials of $0.9V_s$,⁷ but below this, and especially near V_i , these alone would not be effective. Deferring for the present a more detailed consideration of the trigger region, two modes of operation in the main gap may be distinguished, according to polarity.

(5.2) Main-Gap Breakdown Mechanisms

When the main gap is negative, photo-electrons will feed into the advancing low-density region (l.d.r.) and will produce avalanches within it [Fig. 9(a)]. The electrons resulting from the avalanches will ultimately go into the anode, leaving a positive-ion space charge. As the l.d.r. progresses, so will the space charge, until, when the gap has been bridged sufficiently, a field strength is built up ahead of the l.d.r. to permit the development of a positive streamer, as envisaged for non-uniform gaps.⁹ This latter stage would not be governed by the advance of the l.d.r. and might be very rapid. Close to V_s , the time-lag T can be very short, since, even ignoring photo-electric effects, considerable space charges produced in the trigger spark channel could initiate a positive streamer immediately. Below V_i , the advancing l.d.r. peters out before the necessary space-charge field is attained.

If the main gap is positive the photons are used inefficiently; electrons produced ahead of the l.d.r., and especially those at the anode, are lost, since they will move into the anode [Fig. 9(b)], producing few secondary electrons if the potential is well below V_s . The only useful electrons are those produced either in the l.d.r. itself or on the cathode side of it. Provided that a supply of electrons can be maintained (see Section 5.3), electron avalanches will occur in the l.d.r., in which region alone appreciable values

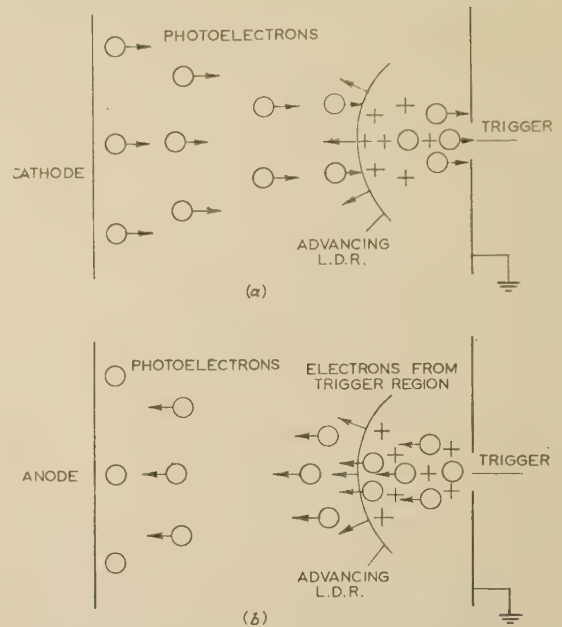


Fig. 9.—Main-gap breakdown mechanisms.

(a) Main gap negative.

(b) Main gap positive.

○ Electrons. + Positive ions.

of α would be possible. Again, the low-mobility positive ions would be left behind to form a space charge, and this would grow forward as the l.d.r. progressed. Behind the l.d.r. front a plasma-like extension of the cathode is possible, provided that an electron source is maintained. After sufficient outward movement in this way, the field between anode and cathode extension may then be such that, together with the electrons produced through the l.d.r., breakdown follows in a normal manner. As V_s is approached, there will again be a shortening of the time lag T due to less of the gap being bridged by the l.d.r. before breakdown, whilst below V_i , breakdown ceases because the l.d.r. peters out, as for the negative gap.

Certain deductions in conformity with experiment can be made already from the discussions above. For both polarities of V , the magnitude of the l.d.r. largely determines V_i , and therefore a distinct correlation between the magnitude of the trigger discharge and V_i is expected, as is shown in the Table. Furthermore, the more efficient use of electrons when V is negative would produce a lower value of V_i than when V is positive. This result is found experimentally in Fig. 5. Both the photo-electrons produced during the trigger discharge and in the subsequent outward-moving l.d.r. are necessary, since a trigger pulse insufficient to cause a discharge does not produce a breakdown in the main gap. The mere potential change alone is not enough to cause a breakdown, even very close to V_s^3 . Near V_i , where T is large, the production of the l.d.r. is the vital factor, and even when the trigger electrode is withdrawn 2 mm (see Section 4.2), V_i is not affected since the l.d.r. can still be produced.

The observed effects with screens (see Section 4.2) are also evidence for the general correctness of the proposed mechanisms. Conditions at the undrilled electrode are comparatively unimportant for both polarities, so that it is not surprising that a paper screen in this position has little observed effect on V_i . Minor effects were probably beyond the limit of measurement. Movement of the screen towards the trigger electrode restricts charge motion and photo-electric action in the gap, and thereby raises V_i . The main actions shown in Fig. 9 also help to explain

the observed behaviour when the screen is close to the trigger. When V is positive, the screen cuts off the cathode extension since the l.d.r. cannot move into the gap and the initial positive-ion space charge associated with the trigger spark is ineffective. Thus no trigger action is obtained. If the polarity is reversed, a very limited triggering range may still exist, since the initial positive-ion space charge mentioned above could still produce field conditions for a streamer process beyond the screen. This triggering range is found to be very small, as might be expected.

Lastly the effect of irradiation is easily understood. With V positive, electrons produced in the bulk of the main gap are lost in any case, and with V negative, the breakdown is mainly dependent on the l.d.r. movement and the existence of an appreciable value of α within it, so that electrons produced by irradiation become insignificant even at the limit, V_I .

(5.3) Trigger Mechanisms

The finer details of these processes are shown by the dependence of V_I and T on the polarity of v in the trigger discharge, as in Figs. 6 and 7. The preceding argument has assumed that the l.d.r. is independent of the trigger polarity, but to explain the reasons for the polarity characteristics shown in these Figures, it seems necessary to assume that, when the main gap ultimately breaks down, it does so between the undrilled electrode and the trigger itself. That the final transfer to the surrounding main electrode occurs very rapidly, and is of secondary importance only, has already been suggested by Wilkinson.⁴ Without some such preference, no obvious reason exists for a polarity effect. The geometrical arrangement of the trigger electrode would encourage such a preference in any case. As discussed earlier, the trigger spark channel becomes the ejected low-density region, but in the neighbourhood of the trigger electrodes, conditions will depend on polarity. Thus, at the trigger anode, an electron space charge left at the cessation of the discharge will rapidly disperse owing to its high mobility, but the corresponding positive-ion sheath at the cathode will exist for longer times. Apart from these space charges, the cathode spot will have become heated, and will be emitting electrons even after the discharge has ceased. Such factors must determine the polarity effects observed. It is unlikely that the polarity of the trigger potential v alone is responsible for the effects, for the following reasons. First, the results cannot be explained by straightforward addition of the potentials V and v , and secondly, the trigger polarity is still effective for the longer time lags near V_I long after the trigger discharge has ceased (see Figs. 6 and 7). Once the trigger discharge commences, the voltage across the trigger electrodes will collapse to a small value and will then be a minor influence on the main-gap events. Thus an explanation is sought in terms of residual space charges at the trigger electrodes immediately after the discharge and electron emission from the hot cathode spot.

Within about 10% of V_s , the short time lags may be attributed to the intense photo-electric action in the gap, in agreement with previous observations.¹⁰ It is below this level that distinct polarity effects occur. First, consider Fig. 7 for V negative. The time lags which, except near V_I , are greatest for v positive, increase steadily as V is decreased. The main mechanism of breakdown envisaged [Fig. 9(a)] involves a positive-ion region at the main-gap anode progressing streamer-like towards the cathode, and this process will be encouraged by an initial positive-ion region at the trigger. This condition obtains when v is negative but not when it is positive, so that the longer time lags for positive trigger polarity are understandable. However, as $V \rightarrow V_I$ there must be a reversal of conditions, since V_I is found to be lower for v positive (Fig. 5). Near this limit, since the voltage is low, the attainment of adequate ionization for the

beginning of a streamer in the deteriorating l.d.r. is difficult, and it is possible that the electron emission from a hot trigger cathode (which has been denuded of the original positive ions during the long time lag) inhibits this further. Conditions can then be more favourable when v is positive because of the absence of electron emission.

It should be remembered that there is considerable lack of precision in determining V_I in all measurements with V negative, so that the above argument cannot be more than tentative. The indefinite nature of V_I probably reflects the statistical nature of streamer propagation and gap breakdown at these low voltages.

For positive values of V below that at which photo-electric action alone could cause breakdown (about $0.9V_s$), there is a rapid increase in T when v is positive but no significant increase when v is negative until V_I is approached (see Fig. 6). If the main mechanism for this case is recalled [see Fig. 9(b)], it is not difficult to give a reasonable explanation. When v is positive, the trigger is a relatively weak source of electrons for ionization in the l.d.r., and this advancing region is not therefore of low conductivity. It will need to advance for an appreciable distance into the gap before a normal breakdown occurs, and there will be a correspondingly large value of T . When v is negative, the emission from the trigger ensures adequate electrons and a small value of T . Near V_I it is suggested that the main-gap field becomes inadequate for sufficient ionization in the l.d.r., and with this is the added necessity that the l.d.r. must advance further into the gap before breakdown, and will therefore be less easily ionized. Whilst there may be a small difference in the values of V_I for positive and negative values of v , it may not have been detectable in the present experiments. If, as seems possible, the field strength in the low-density region is the criterion for determining V_I , no dependence on trigger polarity would be expected.

According to the process outlined above, the l.d.r. must advance well into the gap when both V and v are positive and V is below $0.9V_s$. The time lag T in this case will be almost entirely due to this movement. For a range of gap settings at comparable field strengths, equal proportions of the gaps would need to be traversed by the l.d.r. before breakdown occurred. This is strikingly confirmed for $V = 0.75V_s$ by Fig. 10, which

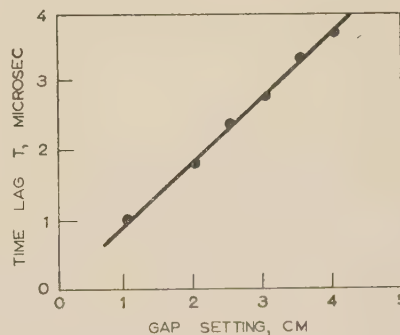


Fig. 10.—Relationship between time lag T and gap setting for positive main gap.

$C = 0.04 \mu\text{F}$.
 $R_2 = 0$.
 $V = 0.75 V_s$
 $v = 10 \text{ kV (positive)}$.

shows a linear relationship between time lag and gap. The velocity of travel of the supposed l.d.r. cannot be precisely estimated, since the travel required before breakdown occurs is not known, but the upper limit resulting from the assumption that the whole gap must be traversed is about 10^6 cm/sec . This is probably a considerable over-estimate, and an actual velocity closer to 10^5 cm/sec is more likely. Velocities of this order have

been observed in the growth of spark channels,⁶ and therefore may correspond to the growth of the trigger spark channel which subsequently becomes the low-density region.

Lastly, the effect of the trigger resistance R_2 on time lag should also be noted. In all the experiments discussed, R_2 has been zero and the trigger discharge rapidly completed before breakdown of the main gap, except, perhaps, when V was near V_s , in which case photo-electric mechanisms predominated. Increase in R_2 will alter the rate of growth of the spark channel, and therefore the l.d.r. velocity, and, by limiting the maximum discharge current, will limit the number of charge carriers also. Increase in T with increase in R_2 , as found for all polarity combinations (see Fig. 8), is thus a logical outcome of the changed trigger conditions. Furthermore, the fact that, when R_2 is zero, variation in C and v (Section 4.3) did not change T , except near V_1 , is also to be expected, since the discharge must be completed very rapidly before the main breakdown occurs.

(6) DIVERTER OPERATION

The essential features of a complete diverter employing the trigatron have been described in Section 2. Such a diverter was tested under direct and impulse voltage conditions up to 300 kV, using the circuits shown in Fig. 11. To protect a spark-gap

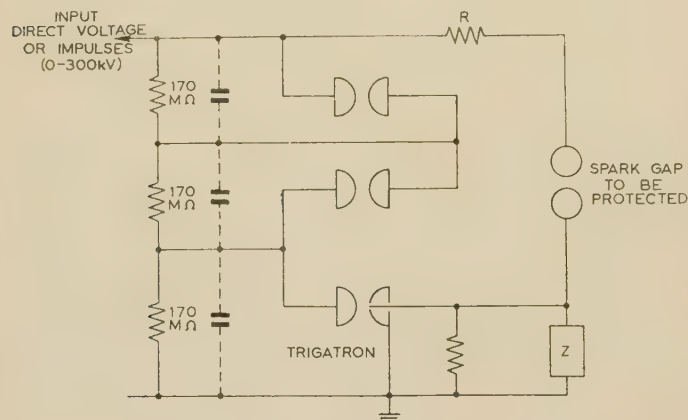


Fig. 11.—Spark-gap diverter circuit for direct and impulse voltages.

For direct voltages, $R = 20 \text{ M}\Omega$ and Z is a suitable inductance. For 1/50 microsec impulse voltages $R = 20 \text{ k}\Omega$, Z is a capacitance of $40 \mu\text{F}$ and the $170 \text{ M}\Omega$ resistors are shunted by $250 \mu\text{F}$ capacitors.

subjected to direct voltage, the gap was earthed through an inductor, the voltage developed across this on breakdown supplying the trigger pulse for the trigatron stage of the diverter. The circuit is simple and effective. The breakdown of a rod-gap was limited to a faint blue discharge, whilst for a 12.5 cm sphere-gap, a luminous filamentary discharge developed without bridging the gap completely. As expected from the trigatron characteristics (Figs. 5 and 7), operation was easier with negative voltage on the spark-gap.

For impulse voltages, the diverter was modified as in Section 2, and the series inductor was replaced by a capacitor of sufficient magnitude to prevent the trigatron operating unless the spark-gap broke down, but not so large that an excessive charging current was caused. A capacitor of approximately $40 \mu\text{F}$ was suitable. Results with this circuit for a 1/50 microsec impulse were similar to those for direct voltage.

To demonstrate the effectiveness of the equipment as a controlled chopping device for impulse waves, a circuit to provide a trigger pulse after a predetermined delay was constructed. Operation of the impulse generator was made to fire a thyatron, which in turn fired a second thyatron after a controlled delay.

A third high-voltage thyatron was then operated by the second to produce the triggering pulse for the diverter. With this circuit, standard 1/50 microsec impulse waves have been chopped on the wavetail at voltages as low as 50% of the impulse peak value of about 300 kV. Provided that the main gap voltage and the trigger pulse were of the same polarity, the instant of chopping was very consistent, as expected from the results of Section 4.3.

(7) CONCLUSIONS

The paper provides a more extensive theory of the trigatron behaviour than hitherto, although it incorporates and agrees with the earlier suggestions.^{3,4,5,12} Further measurements under more accurately controlled conditions are required to confirm these suggested mechanisms, especially those concerned with trigger polarity influence. It is interesting to note, in support of the suggested role of the low-density region, that recent work on a thermally triggered spark-gap¹¹ has shown that breakdown at voltages well below V_s can occur if a hot filament or even a flame is used in place of the trigger discharge. The trigatron does appear to provide a suitable means of studying gas-breakdown processes and especially their time-dependence in gaps which are initially subjected to voltages below the static breakdown voltage. The advantage of being able to influence breakdown in a controlled manner is obvious.

Apart from the employment of the trigatron in gas-breakdown investigations, it seems to be an extremely useful controlled switch for high-voltage circuits and can be made to operate rapidly and consistently over a wide voltage range provided that due attention is paid to the relevant characteristics (see Figs. 5, 6, 7 and 8). The most efficient overall operation is obtained when the main gap has negative voltage V , and a negative trigger pulse v is used, the trigger discharge energy being the maximum that can be conveniently obtained.

Under such conditions it offers a means of efficient spark energy diversion and impulse wave 'chopping' at higher voltages than is possible by other means.

(8) ACKNOWLEDGMENTS

The authors wish to thank Prof. W. J. John in whose laboratories the work was performed, and also their colleagues for considerable help.

One of the authors (A. M. S.) acknowledges a grant from the Royal Norwegian Council for Scientific and Industrial Research, and thanks the Directors of Standard Telefon og Kabelfabrik for leave of absence. Thanks are also due to the University of London for the award of a Turner and Newall Fellowship to the other author (T. J. L.).

(9) REFERENCES

- (1) SAXE, R. F., and LEWIS, T. J.: 'Measurement of Statistical Time Lag of Breakdown in Gases and Liquids', *British Journal of Applied Physics*, 1955, 6, p. 211.
- (2) British Patent No. 578510.
- (3) CRAGGS, J. D., HAINE, M. E., and MEEK, J. M.: 'The Development of Triggered Spark Gaps for High-Power Modulators', *Journal I.E.E.*, 1946, 93, Part IIIA, p. 963.
- (4) WILKINSON, K. J. R.: 'Some Developments in High-Power Modulators for Radar', *ibid.*, 1946, 93, Part IIIA, p. 1090.
- (5) HUSBANDS, A. S., and HIGHAM, J. B.: 'The Controlled Tripping of High-Voltage Impulse Generators', *Journal of Scientific Instruments*, 1951, 28, p. 242.
- (6) MEEK, J. M., and CRAGGS, J. D.: 'Electrical Breakdown in Gases' (Oxford, Clarendon Press, 1953), Chapter 10.

- (7) WHITE, H. J.: 'The Variation of Sparking Potential with Intense Ultraviolet Illumination', *Physical Review*, 1935, **48**, p. 113.
 - (8) BAUER, W. M., and COBINE, J. D.: 'Gap Recovery Strength of A.C. Arcs at High Pressure', *General Electric Review*, 1941, **44**, p. 315.
 - (9) LOEB, L. B., and MEEK, J. M.: 'The Mechanism of the Electric Spark' (Stanford University Press, 1941).
 - (10) WHITE, H. J.: 'Effect of Intense Illumination on Time Lag in Static Spark Breakdown', *Physical Review*, 1936, **49**, p. 507.
 - (11) BROADBENT, T. E., and WOOD, J. K.: 'A Thermally Triggered Spark Gap', *British Journal of Applied Physics*, 1955, **6**, p. 368.
 - (12) HARDY, D. R., and BROADBENT, T. E.: 'The Control of High-Voltage Impulse Generators', *Beama Journal*, 1955, **62**, p. 63.
-

SYNTHESIS OF TRANSMISSION SYSTEMS IN TERMS OF TANDEM-CONNECTED QUADRIPOLES

By P. W. SEYMOUR, Associate Member, and S. DØSSING, M.Sc.(E.E.).

(The paper was first received 20th April, 1954, in revised form 24th February and 25th October, 1955, and in final form 30th January, 1956. It was published as an INSTITUTION MONOGRAPH in August, 1956.)

SUMMARY

A theory for the tandem connection of quadripoles is presented, using as basic parameters the operating factor and input and output impedances of a quadripole when operating under reference conditions between arbitrarily chosen source and load impedances. These parameters are termed the reference parameters, and the source and load impedances are called the nominal impedances. Simple extension leads to the reference loss parameters.

Operating-loss and input-return-loss formulae are derived for the non-reference operation of a quadripole, and, in conjunction with return-loss formulae obtained for quadripole junctions, are used to determine overall operating-loss characteristics of a transmission system when quadripoles are added or removed. Furthermore, these formulae are used for transmission-system synthesis.

A technique is developed which permits convenient and precise calculation of return and operating loss, or alternatively enables calculation of their limits.

The formulae derived facilitate the rapid solution of a variety of practical transmission problems, and their application in some instances is illustrated by the inclusion of numerical examples.

LIST OF PRINCIPAL SYMBOLS

V = Complex voltage and e.m.f.

I = Complex current.

n = An integer.

$G = g e^{j\phi}$ = Reference operating factor.

OL = Reference operating loss, $20 \log_{10} g$, decibels.

Z = Impedance. (Various subscripts are attached.)

Z_{N1}, Z_{N2}, Z_N , etc. = Nominal impedances.

Z_S = Source impedance.

Z_L = Load impedance.

Z_{R1} = Reference input impedance.

Z_{R2} = Reference output impedance.

A_{R1}, A_{R2} = Reference inverse reflection coefficients, corresponding to Z_{R1} and Z_{R2} , respectively.

A_S, A_L = Inverse reflection coefficients, corresponding to Z_S and Z_L , respectively. (In general $A = |A| e^{j\theta} = a e^{j\theta}$.)

Z_1, Z_r = Dissimilar junction impedances.

A_J = Junction inverse reflection coefficient.

$$(A_J = \frac{Z_r + Z_1}{Z_r - Z_1})$$

K, X = Complex quantities defined in text.

$k = |K|$ (Subscripts 'max' and 'min' may be attached.)

VL = Variation loss, $20 \log_{10} k$ (dB). Subscripts 'max' and 'min' may be attached.

$G_{EE,0}, G'_{EE,0}$ = Operating factors, respectively, of an ideal transducer (ratio Z_{N2}/Z_{N1}) and an interacting quadripole between Z_S and Z_L .

G_{EE}, G'_{EE} = Overall reference (or non-reference) operating factors of a transmission system before and after a substitution operation, respectively.

$G_{EE,s}, G'_{EE,s}$ = Overall reference operating factors of transmission systems synthesized from two and three quadripoles, respectively.

$OL_{EE,0}$, etc. = $20 \log_{10} |G_{EE,0}|$, etc.

$J(f)$ = Frequency response. (Subscripts may be attached as defined in the text.)

$J_1, J_2, J_{EE,1}, J_{EE,2}$ = Frequency-response limits.

f_0 = Reference frequency of a channel.

f_0, n = Reference frequency of n th channel in a group.

f_G = Group reference frequency.

RL = Return loss, decibels.

$= 20 \log_{10} |A| = 20 \log_{10} a$.

(Subscripts for RL, a and A correspond. They are defined in the text.)

Z_{in}, A_{in} = Input impedance and inverse reflection coefficient of interacting quadripole under non-reference conditions.

$$A_{in} = \frac{Z_{in} + Z_{N1}}{Z_{in} - Z_{N1}}$$

$G_i = g_i e^{j\phi_i}$ = Image reference operating factor.

(1) INTRODUCTION

The performance of a quadripole may, in practice, be specified in terms of its operating factor and its input and output impedances when operating between arbitrarily chosen source and load impedances.

This method of specification can be used to ensure reasonable matching conditions when quadripoles are tandem-connected, and it further possesses the advantage that the parameters used may be conveniently measured.

A complete theoretical treatment based on these parameters does not appear to have been previously published. The authors have, however, applied them to a special case of tandem connection,^{1,6} and encouraged by the simple results obtained, they developed the general theory presented in the paper.

The parameters mentioned above are termed the reference parameters, as distinct from the image or iterative parameters. The reference parameters depend on the particular source and load impedances chosen. The latter are termed the nominal source and load impedances, or simply the nominal impedances.

The nominal impedances are often of zero angle, and in many cases they correspond closely to the image or design impedances of the quadripole.

In order to obtain generalized results, two distinct classes of quadripole are considered. The first, or interacting class, is comprised of all linear passive four-terminal networks, such as filters, transformers, attenuators, equalizers, lengths of cable, etc. The second, or non-interacting class, contains quadripoles whose input and output impedances do not vary with varying terminating impedances. Quadripoles of this type are frequently

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Seymour is in the Radar and Telecommunications Section, Department of Defence Production, Australia, and was formerly in the Office of the Engineer-in-Chief, Postmaster-General's Department, Australia.

Mr. Døssing is in the Office of the Engineer-in-Chief, Postmaster-General's Department, Australia.

encountered in practice, and comprise, with a few exceptions, equipments incorporating electron-tube amplifiers.

Other equipments, such as the modulators encountered in carrier telephone equipment, are of a more complex nature and do not belong to either of the classes. The formulae obtained may, however, often be applied to such equipment with good approximation (see Section 7.1).

(2) OPERATING FACTOR AND OPERATING LOSS

Frequently a number of different nominal impedances are encountered within a transmission system, and in such cases the concept of operating factor² is more useful than that of insertion factor, since the former, unlike the latter, combines the actual loss within a quadripole with its impedance-matching properties. The symbol Z_N , with appropriately numbered subscript, will be used for nominal impedances throughout the paper.

The operating factor, G , of a quadripole Q operating between a source impedance Z_{N1} and a load impedance Z_{N2} , is defined as the ratio obtained by dividing the current I_0 , which would flow in Z_{N2} if it had been connected via an ideal impedance-matching transducer of impedance ratio Z_{N2}/Z_{N1} to the source, by the current I_1 , which would flow in Z_{N2} upon replacement of

For calculation purposes the ideal transducer may be considered as an abstract ideal transformer having a turns ratio $\sqrt{Z_{N2}/Z_{N1}}$. This turns ratio becomes real if Z_{N2}/Z_{N1} is real and positive. If $Z_{N1} = Z_{N2}$ the operating and insertion factors become identical.

If the modulus and angle of G are represented by g and ϕ , respectively, the operating loss, OL , is defined as $20 \log_{10} g$ decibels.

(3) REFLECTION COEFFICIENT AND RETURN LOSS

In accordance with the classical quadripole theory, reflections are said to occur at a junction of dissimilar impedances. If the impedances at such a junction are Z and Z_N , the reflection coefficient may be expressed as

$$\frac{Z - Z_N}{Z + Z_N}$$

It is often more convenient to analyse in terms of the inverse reflection coefficient

$$A = a e^{j\theta} = \frac{Z + Z_N}{Z - Z_N}, \text{ where } a = |A|$$

The return loss, RL , at this impedance junction may be defined as $20 \log_{10} a$ decibels.

It can be simply shown that the locus of Z/Z_N for a constant value of a is a circle of radius $2a/(a^2 - 1)$ and centre at $(a^2 + 1)/(a^2 - 1), 0$.

Similarly, the locus of Z/Z_N for a constant value of θ is a circle of radius $\csc \theta$ and centre at $(0, -\cot \theta)$.

These results have been used to prepare Fig. 2, from which the return loss (in decibels) and θ can be obtained for given values of the complex variable $Z/Z_N = r + jx$.

The return loss may be regarded as a measure of the inequality existing between Z and Z_N , and since it may be measured directly with simple instruments, the specification of the minimum limit of return loss to be permitted between Z and Z_N is a convenient method of expressing the tolerance of Z relative to the desired value Z_N .

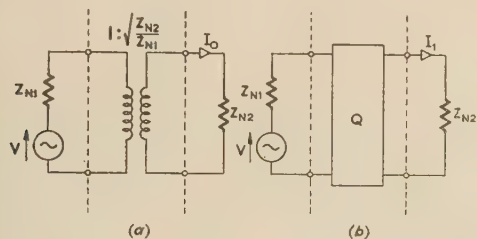


Fig. 1.—Operating factor of an interacting quadripole.

- (a) Ideal transducer between Z_{N1} and Z_{N2} .
(b) Interacting quadripole between Z_{N1} and Z_{N2} .

the ideal transducer by the quadripole Q . Thus G equals I_0/I_1 , as illustrated in Fig. 1.

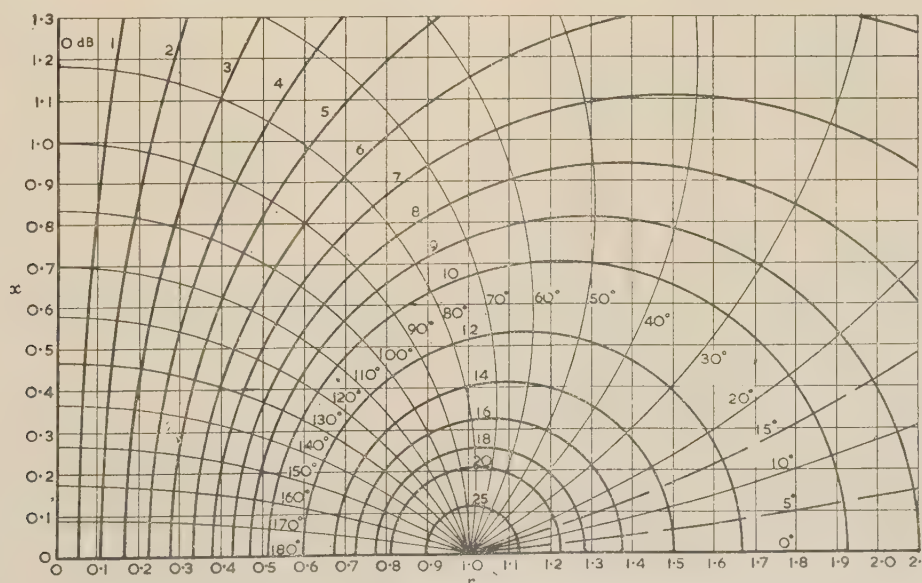


Fig. 2.—Graphs of RL , in decibels, and $\angle \theta$ in terms of $Z/Z_N = r + jx$.

θ is negative for $r + jx$ and positive for $r - jx$.

(4) DEFINITIONS OF REFERENCE CONDITIONS AND QUADRIPOLE REFERENCE PARAMETERS

When a quadripole is operating between its nominal impedances it is said to be operating under reference conditions, and conversely it is said to be operating under non-reference conditions when operating between any other impedances.

At junctions of quadripoles in tandem it is assumed that the nominal load impedance of the driving quadripole is equal to the nominal source impedance of the driven quadripole, and this common impedance is termed the 'nominal junction impedance'.

For convenience, where there is no possibility of ambiguity, the term 'nominal impedance' may be used to cover any of the three terms 'nominal source impedance', 'nominal load impedance' and 'nominal junction impedance'.

A precise definition of the reference impedances may now be obtained from Fig. 3. As shown, the quadripole input and out-

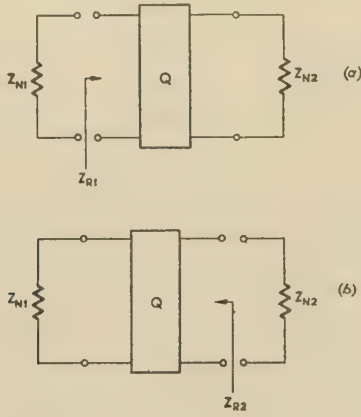


Fig. 3.—Definition of quadripole reference impedances.

Nominal impedances Z_{N1} and Z_{N2} .

$$(a) \text{ Input inverse reflection coefficient } A_{R1} = \frac{Z_{R1} + Z_{N1}}{Z_{R1} - Z_{N1}}$$

$$\text{Reference input return loss } RL_{R1} = 20 \log_{10} a_{R1}$$

$$(b) \text{ Output inverse reflection coefficient } A_{R2} = \frac{Z_{R2} + Z_{N2}}{Z_{R2} - Z_{N2}}$$

$$\text{Reference output return loss } RL_{R2} = 20 \log a_{R2}$$

put impedances Z_{R1} and Z_{R2} under reference conditions may be referred to, respectively, as the 'reference input impedance' and the 'reference output impedance', while in a similar manner the return losses, RL_{R1} and RL_{R2} , may be referred to, respectively, as the 'reference input return loss' and the 'reference output return loss'.

Similarly the operating factor, G , and the operating loss, OL , under reference conditions, as shown in Fig. 1, will be referred to, respectively, as the 'reference operating factor' and the 'reference operating loss'.

The reference return losses and the reference operating loss may be collectively referred to as the 'reference loss parameters'.

The reference return losses will, for all practical applications envisaged, be positive, and this restriction is assumed throughout, although most of the formulae developed will also apply where negative reference return losses are encountered.

(5) TRANSMISSION SYSTEMS COMPOSED OF TANDEM-CONNECTED QUADRIPOLES

In a transmission system a quadripole linking two junctions having different nominal impedances Z_{N1} and Z_{N2} should not be removed, as this would result in abnormal impedance-matching conditions at the newly formed junction. Replacement with an

ideal impedance-matching transducer is permissible, although of little practical use. However, the mathematical treatment given in Sections 15.1–15.3 for this case may simply be extended to cover replacement of one quadripole with another. Further, if $Z_{N1} = Z_{N2}$, so that the ideal transducer becomes of unit ratio and acts as a straight-through connection, the results obtained cover both the addition and removal of quadripoles.

(5.1) Overall Operating Factor

Consider a transmission system as shown in Fig. 4(a). It is desired to replace the ideal transducer of impedance ratio

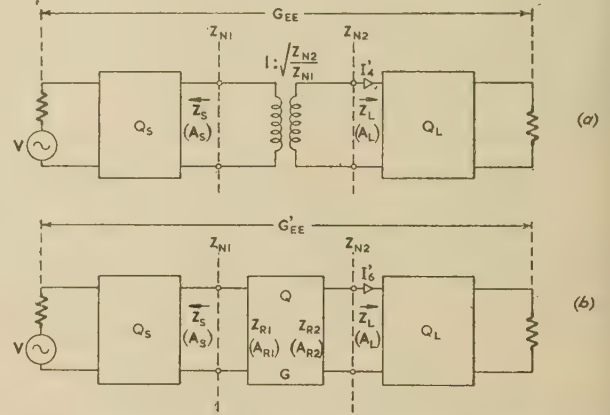


Fig. 4.—Replacement of an ideal transducer with a quadripole.

Z_{N2}/Z_{N1} with a practical quadripole Q having reference parameters Z_{R1} , Z_{R2} and G , as shown in Fig. 4(b). If the original and new overall operating factors (either reference or non-reference) are G_{EE} and G'_{EE} , respectively (the subscript EE indicating end-to-end), using the nomenclature shown it is seen that

$$G'_{EE} = G_{EE} \frac{I_4}{I_6}$$

The ratio I_4/I_6 may be expressed in terms of Z_{N1} , Z_{N2} , Z_S , Z_L and the quadripole parameters as shown in Sections 15.1–15.3. Thus, for a quadripole of the interacting type

$$G'_{EE} = G_{EE} GK$$

where

$$K = \frac{\left(\frac{Z_S}{Z_{N1}} - 1\right)\left(\frac{Z_L}{Z_{N2}} - 1\right)}{2\left(\frac{Z_S}{Z_{N1}} + \frac{Z_L}{Z_{N2}}\right)}$$

$$\left[\frac{2\left(\frac{Z_S}{Z_{N1}} + \frac{Z_{R1}}{Z_{N1}}\right)}{\left(\frac{Z_S}{Z_{N1}} - 1\right)\left(\frac{Z_{R1}}{Z_{N1}} + 1\right)} \frac{2\left(\frac{Z_{R2}}{Z_{N2}} + \frac{Z_L}{Z_{N2}}\right)}{\left(\frac{Z_{R2}}{Z_{N2}} + 1\right)\left(\frac{Z_L}{Z_{N2}} - 1\right)} - \frac{1}{G^2} \right]$$

Introducing the inverse reflection coefficients,

$$A_S = \frac{\frac{Z_S}{Z_{N1}} + 1}{\frac{Z_S}{Z_{N1}} - 1}, \quad A_{R1} = \frac{\frac{Z_{R1}}{Z_{N1}} + 1}{\frac{Z_{R1}}{Z_{N1}} - 1},$$

$$A_{R2} = \frac{\frac{Z_{R2}}{Z_{N2}} + 1}{\frac{Z_{R2}}{Z_{N2}} - 1} \quad \text{and} \quad A_L = \frac{\frac{Z_L}{Z_{N2}} + 1}{\frac{Z_L}{Z_{N2}} - 1}$$

and using the expressions for impedance ratios derived from these,

$$K = \frac{A_S A_L}{A_S A_L - 1} \left(\frac{A_S A_{R1} - 1}{A_S A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L} - \frac{1}{A_S A_L G^2} \right)$$

This is conveniently written as

$$K = \frac{A_S A_L}{A_S A_L - 1} X \frac{X A_S A_L G^2 - 1}{X A_S A_L G^2}$$

where

$$X = \frac{A_S A_{R1} - 1}{A_S A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}$$

Simple extension of these results relating to replacement of an ideal transducer with a quadripole leads to results for replacement of one quadripole with another. Thus if the parameters associated with an original quadripole Q_A are identified by the

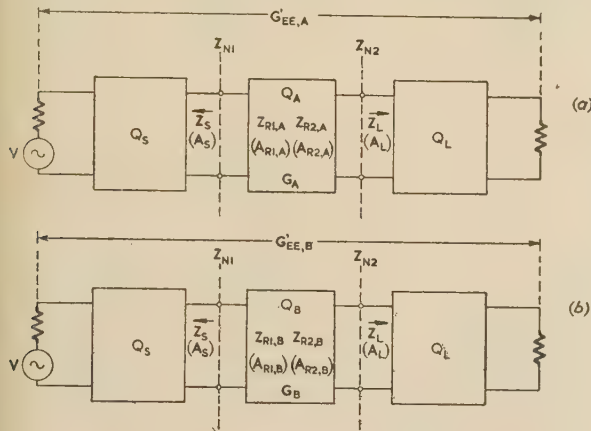


Fig. 5.—Replacement of quadripole Q_A with quadripole Q_B .

additional subscript A , and those of a replacing quadripole Q_B by an additional subscript B , as shown in Fig. 5, it is readily seen that

$$G'_{EE,B} = G'_{EE,A} \frac{G_B}{G_A} K$$

where

$$K = \frac{\frac{A_S A_{R1,B} - 1}{A_S A_{R1,B}} \frac{A_{R2,B} A_L - 1}{A_{R2,B} A_L} - \frac{1}{A_S A_L G_B^2}}{\frac{A_S A_{R1,A} - 1}{A_S A_{R1,A}} \frac{A_{R2,A} A_L - 1}{A_{R2,A} A_L} - \frac{1}{A_S A_L G_A^2}} = \frac{X_B \frac{X_B A_S A_L G_B^2 - 1}{X_B A_S A_L G_B^2}}{X_A \frac{X_A A_S A_L G_A^2 - 1}{X_A A_S A_L G_A^2}}$$

and

$$X_B = \frac{A_S A_{R1,B} - 1}{A_S A_{R1,B}} \frac{A_{R2,B} A_L - 1}{A_{R2,B} A_L}$$

$$X_A = \frac{A_S A_{R1,A} - 1}{A_S A_{R1,A}} \frac{A_{R2,A} A_L - 1}{A_{R2,A} A_L}$$

The terms of the form $(X A_S A_L G^2 - 1)/(X A_S A_L G^2)$ become unity if $|A_S|$, $|A_L|$ or $|G|$ becomes infinite, and may therefore be considered to represent an interaction phenomenon.

The formulae developed apply to replacements with interacting quadripoles. From these formulae may be obtained the solution of certain special cases.

The first case for consideration is that of replacement of non-interacting quadripoles. In this case the interaction terms vanish and the formulae assume the form shown in Table 1.

The second case is that of replacement of both types of quadripole when $Z_S = Z_{N1}$. The relevant formulae, shown in Table 1, are of particular use when performing replacements at the sending end of a transmission system, where, for most test purposes, the source impedance is chosen to be equal to the nominal im-

pedance. Similar formulae apply at the receiving end of a system for $Z_L = Z_{N2}$.

The third case is that of replacement of quadripoles having known image reference parameters. A quadripole operating under image-matched reference conditions has image reference impedances (image impedances) $Z_{R1} = Z_{N1}$ and $Z_{R2} = Z_{N2}$ (so that $|A_{R1}| = |A_{R2}| = \infty$) and an image reference operating factor G_i . The image reference operating loss, $OL_i = 20 \log_{10} g_i$ decibels, may be identified as the image attenuation coefficient expressed in decibels. The formulae for this case are also shown in Table 1.

It is seen from Table 1 that K is composed of factors of the form $X/(X - 1)$ and $(X - 1)/X$. This also applies to practically all other formulae derived later. If all complex quantities are

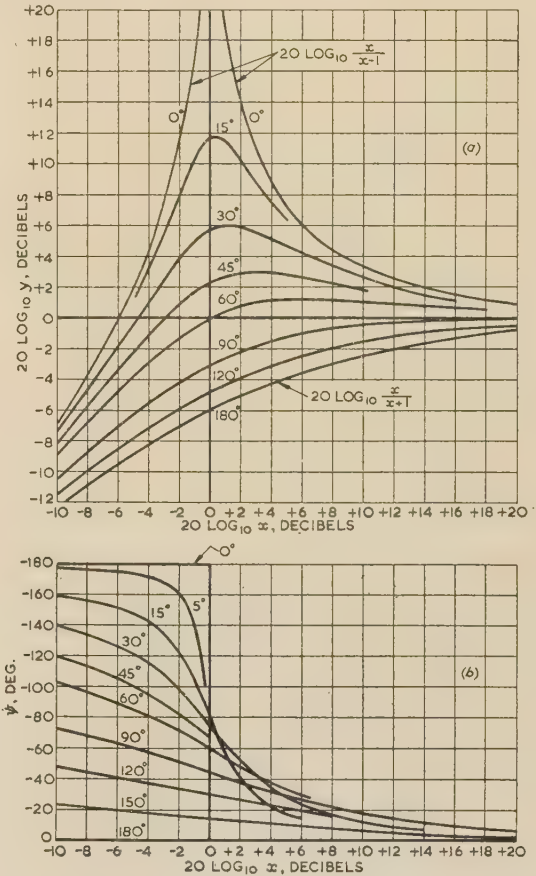


Fig. 6.—The universal function $y e^{j\theta} = \frac{x e^{j\theta} - 1}{x e^{j\theta} + 1}$

(a) $20 \log_{10} y$ versus $20 \log_{10} x$ (parameter, θ).
(b) θ versus $20 \log_{10} x$ (parameter, θ).

given in terms of $20 \log_{10} |X|$ decibels and the angle of X , numerical calculations may conveniently be carried out by means of Tables giving $20 \log_{10} |X/(X - 1)|$ and the angle of $X/(X - 1)$ as functions of $20 \log_{10} |X|$ and the angle of X .

None of the existing indexes to mathematical Tables contains any references to suitable Tables of this type, and the authors have therefore prepared appropriate Tables.* These Tables are sufficiently detailed for most practical applications, and if they are plotted as large-scale graphs in the manner indicated in Fig. 6,

* These Tables may also conveniently be used for calculations involving the feedback formulae

$$\frac{\mu}{1 - \mu\beta} = \frac{1}{\beta} \frac{\mu\beta}{1 - \mu\beta}$$

if the magnitudes of μ and β are expressed in decibels, and the angles of μ and β are known.

Copies of these Tables may be obtained upon application to The Institution.

Table 1

VALUES OF K FOR VARIOUS SUBSTITUTION CASES
Limits of $k = |K|$ are given in Table 2.

	Replacing ideal transducer with quadripole as in Fig. 4 $G_{EE} = G_{EE}G_K$	Replacing quadripole Q_A with quadripole Q_B as in Fig. 5 $G_{EE,B} = G_{EE,A} \frac{G_B}{G_A} K$
Definitions of auxiliary parameters used in this Table	$X = \frac{A_S A_{R1} - 1}{A_S A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}$	$X_A = \frac{A_S A_{R1,A} - 1}{A_S A_{R1,A}} \frac{A_{R2,A} A_L - 1}{A_{R2,A} A_L}$ $X_B = \frac{A_S A_{R1,B} - 1}{A_S A_{R1,B}} \frac{A_{R2,B} A_L - 1}{A_{R2,B} A_L}$
Interacting case	$K = \frac{A_S A_L}{A_S A_L - 1} X \frac{X A_S A_L G^2 - 1}{X A_S A_L G^2}$	$K = \frac{X_B}{X_A} \frac{X_B A_S A_L G_B^2 - 1}{X_B A_S A_L G_B^2} \frac{X_A A_S A_L G_A^2}{X_A A_S A_L G_A^2 - 1}$
Non-interacting case	$K = \frac{A_S A_L}{A_S A_L - 1} X$	$K = \frac{X_B}{X_A}$
$Z_S = Z_{N1}$, or $ A_S = \infty$	$K = \frac{A_{R2} A_L - 1}{A_{R2} A_L}$	$K = \frac{A_{R2,B} A_L - 1}{A_{R2,B} A_L} \frac{A_{R2,A} A_L}{A_{R2,A} A_L - 1}$
$Z_L = Z_{N2}$, or $ A_L = \infty$	$K = \frac{A_S A_{R1} - 1}{A_S A_{R1}}$	$K = \frac{A_S A_{R1,B} - 1}{A_S A_{R1,B}} \frac{A_S A_{R1,A}}{A_S A_{R1,A} - 1}$
Interacting case: $Z_{R1} = Z_{N1}$, $Z_{R2} = Z_{N2}$, or $ A_{R1} = A_{R2} = \infty$ Reference operation = Image operation, i.e. $G = G_i$	$K = \frac{A_S A_L}{A_S A_L - 1} \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2}$	$K = \frac{A_S A_L G_{i,B}^2 - 1}{A_S A_L G_{i,B}^2} \frac{A_S A_L G_{i,A}^2}{A_S A_L G_{i,A}^2 - 1}$
Non-interacting case: $Z_{R1} = Z_{N1}$, $Z_{R2} = Z_{N2}$, or $ A_{R1} = A_{R2} = \infty$	$K = \frac{A_S A_L}{A_S A_L - 1}$	$K = 1$

numerical calculations of quadripole performance, using the formulae presented in the paper, may be carried out more rapidly than by using any other known method.

(5.1.1) Short Length of Voice-Frequency Cable.

The characteristic impedance and propagation coefficient of a 10 lb/mile cable at 3400 c/s are given as

$$Z_0 = 254.2 - j224.4 \text{ ohms}$$

$$\gamma = 3.0 \text{ dB/mile, } \angle 22^\circ \text{ per mile.}$$

It is desired to insert 1 mile of this cable into a transmission system at a point where the junction impedances are both 600 ohms. The effect of this on the overall operating loss is to be determined.

First, an ideal transducer of unity ratio is inserted. This will not affect the overall operating loss. For the purpose of the calculation the nominal impedances Z_{N1} and Z_{N2} at the input and output of the transducer are chosen to be equal to the characteristic impedance of the cable, i.e.

$$Z_{N1} = Z_{N2} = Z_{R1} = Z_{R2} = Z_0 = 254.2 - j224.4 \text{ ohms}$$

Consequently $20 \log_{10} |G_i|$ decibels = 3.0 dB and the image reference operating angle ϕ_i is 22° .

The K -factor applicable to this replacement is given in Table 1. The numerical calculations may proceed as follows:

$$\frac{Z_S}{Z_{N1}} = \frac{Z_L}{Z_{N2}} = 1.33 + j1.17$$

and from Fig. 2

$$20 \log_{10} |A_S| = 20 \log_{10} |A_L| = 6.6 \text{ dB}$$

and

$$\theta_S = \theta_L = -48^\circ$$

Further

$$20 \log_{10} |A_S A_L| = 13.2 \text{ dB, } \theta_S + \theta_L = -96^\circ$$

$$20 \log_{10} |A_S A_L G_i^2| = 19.2 \text{ dB, } \theta_S + \theta_L + 2\phi_i = -52^\circ$$

The components of the K -factor may now be determined by means of graphs as shown in Fig. 6, or from Tables, as follows:

$$20 \log |K| = 20 \log_{10} \left| \frac{A_S A_L}{A_S A_L - 1} \right| + 20 \log_{10} \left| \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \right| = -0.35 - 0.55 = -0.9 \text{ dB}$$

$$\text{Thus: } OL'_{EE} = OL_{EE} + 3.0 - 0.9 = OL_{EE} + 2.1 \text{ dB}$$

It is seen that, at a single frequency, calculations of operating loss for cables of different length may be carried out with little additional effort, since the term $20 \log_{10} |A_S A_L| / (A_S A_L - 1)$ remains the same for all cable lengths.

In examples such as this, a saving in labour is achieved by working in terms of the inverse impedance ratio Z_N/Z rather than Z/Z_N . This is particularly true if a number of calculations of the type shown here are to be made to obtain a frequency

Table 2

LIMITS k_{max} AND k_{min} OF $k = 1K1$

	Replacing ideal transducer with quadripole as in Fig. 4	Replacing quadripole Q_A with quadripole Q_B , as in Fig. 5
Definitions of auxiliary parameters used in this Table	$x_{max} = \frac{a_S a_{R1} + 1}{a_S a_{R1}} \frac{a_{R2} a_L + 1}{a_{R2} a_L}$ <p>for $\theta_S + \theta_{R1} = (2n_1 + 1)\pi$ and $\theta_{R2} + \theta_L = (2n_2 + 1)\pi$ } (1)</p> $x_{min} = \frac{a_S a_{R1} - 1}{a_S a_{R1}} \frac{a_{R2} a_L - 1}{a_{R2} a_L}$ <p>for $\theta_S + \theta_{R1} = 2n_1\pi$ and $\theta_{R2} + \theta_L = 2n_2\pi$ } (2)</p>	$x_{A,max} = \frac{a_S a_{R1,A} + 1}{a_S a_{R1,A}} \frac{a_{R2,A} a_L + 1}{a_{R2,A} a_L}$ <p>for $\theta_S + \theta_{R1,A} = (2n_1 + 1)\pi$ $\theta_{R2,A} + \theta_L = (2n_2 + 1)\pi$ } (1)</p> $x_{B,max} = \frac{a_S a_{R1,B} + 1}{a_S a_{R1,B}} \frac{a_{R2,B} a_L + 1}{a_{R2,B} a_L}$ <p>for $\theta_S + \theta_{R1,B} = (2n_1 + 1)\pi$ $\theta_{R2,B} + \theta_L = (2n_2 + 1)\pi$ } (2)</p> $x_{A,min} = \frac{a_S a_{R1,A} - 1}{a_S a_{R1,A}} \frac{a_{R2,A} a_L - 1}{a_{R2,A} a_L}$ <p>for $\theta_S + \theta_{R1,A} = 2n_1\pi$ $\theta_{R2,A} + \theta_L = 2n_2\pi$ } (3)</p> $x_{B,min} = \frac{a_S a_{R1,B} - 1}{a_S a_{R1,B}} \frac{a_{R2,B} a_L - 1}{a_{R2,B} a_L}$ <p>for $\theta_S + \theta_{R1,B} = 2n_1\pi$ $\theta_{R2,B} + \theta_L = 2n_2\pi$ } (4)</p>
Interacting	$k_{max} = \frac{a_S a_L}{a_S a_L - 1} x_{max} \frac{x_{max} a_S a_L g^2 + 1}{x_{max} a_S a_L g^2}$ <p>if eqn. (1) is met, and in addition</p> $\left. \begin{array}{l} \theta_S + \theta_L = 2n_3\pi \\ \theta_S + \theta_L + 2\phi = (2n_4 + 1)\pi \end{array} \right\} \text{or} \left\{ \begin{array}{l} \theta_S + \theta_L = 2n_3\pi \\ \phi = (n_4 - n_3)\pi + \frac{\pi}{2} \end{array} \right.$	$k_{max} = \frac{x_{B,max}}{x_{A,min}} \frac{x_{B,max} a_S a_L g_B^2 + 1}{x_{B,max} a_S a_L g_B^2} \frac{x_{A,min} a_S a_L g_A^2}{x_{A,min} a_S a_L g_A^2 - 1}$ <p>if eqns. (2) and (3) are met, and in addition</p> $\begin{aligned} \theta_S + \theta_L + 2\phi_A &= 2n_3\pi \\ \theta_S + \theta_L + 2\phi_B &= (2n_4 + 1)\pi \end{aligned}$ <p>leading to $\phi_B - \phi_A = (n_4 - n_3)\pi + \frac{\pi}{2}$</p>
	$k_{min} = \frac{a_S a_L}{a_S a_L + 1} x_{min} \frac{x_{min} a_S a_L g^2 - 1}{x_{min} a_S a_L g^2}$ <p>if eqn. (2) is met, and in addition</p> $\left. \begin{array}{l} \theta_S + \theta_L = (2n_3 + 1)\pi \\ \theta_S + \theta_L + 2\phi = 2n_4\pi \end{array} \right\} \text{or} \left\{ \begin{array}{l} \theta_S + \theta_L = (2n_3 + 1)\pi \\ \phi = (n_4 - n_3)\pi - \frac{\pi}{2} \end{array} \right.$	$k_{min} = \frac{x_{B,min}}{x_{A,max}} \frac{x_{B,min} a_S a_L g_B^2 - 1}{x_{B,min} a_S a_L g_B^2} \frac{x_{A,max} a_S a_L g_A^2}{x_{A,max} a_S a_L g_A^2 + 1}$ <p>if eqns. (1) and (4) are met, and in addition</p> $\begin{aligned} \theta_S + \theta_L + 2\phi_A &= (2n_3 + 1)\pi \\ \theta_S + \theta_L + 2\phi_B &= 2n_4\pi \end{aligned}$ <p>leading to $\phi_B - \phi_A = (n_4 - n_3)\pi - \frac{\pi}{2}$</p>
Non-interacting	$k_{max} = \frac{a_S a_L}{a_S a_L - 1} x_{max}$ <p>if eqn. (1) is met, and $\theta_S + \theta_L = 2n_3\pi$</p>	$k_{max} = \frac{x_{B,max}}{x_{A,min}}$ <p>if eqns. (2) and (3) are met</p>
	$k_{min} = \frac{a_S a_L}{a_S a_L + 1} x_{min}$ <p>if eqn. (2) is met, and $\theta_S + \theta_L = (2n_3 + 1)\pi$</p>	$k_{min} = \frac{x_{B,min}}{x_{A,max}}$ <p>if eqns. (1) and (4) are met</p>
$Z_S = Z_{N1}$, or $ A_S = \infty$	$k_{max} = \frac{a_{R2} a_L + 1}{a_{R2} a_L}$ <p>if $\theta_L + \theta_{R2} = (2n_1 + 1)\pi$</p>	$k_{max} = \frac{a_{R2,B} a_L + 1}{a_{R2,B} a_L} \frac{a_{R2,A} a_L}{a_{R2,A} a_L - 1}$ <p>if $\theta_L + \theta_{R2,A} = 2n_1\pi$ $\theta_L + \theta_{R2,B} = (2n_2 + 1)\pi$</p>
	$k_{min} = \frac{a_{R2} a_L - 1}{a_{R2} a_L}$ <p>if $\theta_L + \theta_{R2} = 2n_1\pi$</p>	$k_{min} = \frac{a_{R2,B} a_L - 1}{a_{R2,B} a_L} \frac{a_{R2,A} a_L}{a_{R2,A} a_L + 1}$ <p>if $\theta_L + \theta_{R2,A} = (2n_1 + 1)\pi$ $\theta_L + \theta_{R2,B} = 2n_2\pi$</p>

[Continued overleaf]

Table 2—continued

LIMITS k_{max} AND k_{min} OF $k = 1K1$

	Replacing ideal transducer with quadripole as in Fig. 4	Replacing quadripole Q_A with quadripole Q_B , as in Fig. 5
$Z_L = Z_{N2}$, or $ A_L = \infty$	$k_{max} = \frac{a_S a_{R1} + 1}{a_S a_{R1}}$ if $\theta_S + \theta_{R1} = (2n_1 + 1)\pi$	$k_{max} = \frac{a_S a_{R1,B} + 1}{a_S a_{R1,B}} \frac{a_S a_{R1,A}}{a_S a_{R1,A} - 1}$ if $\begin{cases} \theta_S + \theta_{R1,A} = 2n_1\pi \\ \theta_S + \theta_{R1,B} = (2n_2 + 1)\pi \end{cases}$
	$k_{min} = \frac{a_S a_{R1} - 1}{a_S a_{R1}}$ if $\theta_S + \theta_{R1} = 2n_1\pi$	$k_{min} = \frac{a_S a_{R1,B} - 1}{a_S a_{R1,B}} \frac{a_S a_{R1,A}}{a_S a_{R1,A} + 1}$ if $\begin{cases} \theta_S + \theta_{R1,A} = (2n_1 + 1)\pi \\ \theta_S + \theta_{R1,B} = 2n_2\pi \end{cases}$
Interacting $Z_{R1} = Z_{N1}$, $Z_{R2} = Z_{N2}$ or $ A_{R1} = A_{R2} = \infty$ and $G_i = g_i e^{j\theta_i}$ Reference operation = Image operation	$k_{max} = \frac{a_S a_L}{a_S a_L - 1} \frac{a_S a_L g_i^2 + 1}{a_S a_L g_i^2}$ if $\begin{cases} \theta_S + \theta_L = 2n_1\pi \\ \phi_i = n_2\pi + \frac{\pi}{2} \end{cases}$	$k_{max} = \frac{a_S a_L g_{i,B}^2 + 1}{a_S a_L g_{i,B}^2} \frac{a_S a_L g_{i,A}^2}{a_S a_L g_{i,A}^2 - 1}$ if $\begin{cases} \theta_S + \theta_L + 2\phi_{i,A} = 2n_1\pi \\ \theta_S + \theta_L + 2\phi_{i,B} = (2n_2 + 1)\pi \end{cases}$ from which $\phi_{i,B} - \phi_{i,A} = (n_2 - n_1)\pi + \frac{\pi}{2}$
	$k_{min} = \frac{a_S a_L}{a_S a_L + 1} \frac{a_S a_L g_i^2 - 1}{a_S a_L g_i^2}$ if $\begin{cases} \theta_S + \theta_L = (2n_1 + 1)\pi \\ \phi_i = n_2\pi - \frac{\pi}{2} \end{cases}$	$k_{min} = \frac{a_S a_L g_{i,B}^2 - 1}{a_S a_L g_{i,B}^2} \frac{a_S a_L g_{i,A}^2}{a_S a_L g_{i,A}^2 + 1}$ if $\begin{cases} \theta_S + \theta_L + 2\phi_{i,A} = (2n_1 + 1)\pi \\ \theta_S + \theta_L + 2\phi_{i,B} = 2n_2\pi \end{cases}$ from which $\phi_{i,B} - \phi_{i,A} = (n_2 - n_1)\pi - \frac{\pi}{2}$
Non-interacting $Z_{R1} = Z_{N1}$, $Z_{R2} = Z_{N2}$ or $ A_{R1} = A_{R2} = \infty$	$k_{max} = \frac{a_S a_L}{a_S a_L - 1}$ if $\theta_S + \theta_L = 2n_1\pi$	$k_{max} = 1$
	$k_{min} = \frac{a_S a_L}{a_S a_L + 1}$ if $\theta_S + \theta_L = (2n_1 + 1)\pi$	$k_{min} = 1$

response characteristic. Thus, when working with the inverse impedance ratio, it is convenient to write

$$A_S = -\frac{\frac{Z_{N1}}{Z_S} + 1}{\frac{Z_{N1}}{Z_S} - 1}$$

By use of the curves of Fig. 2, $20 \log_{10} |A_S|$ decibels is obtained as before, while the magnitude and sign of the angle is $(\theta_S - 180^\circ)$, where θ_S is read directly from the curves.

(5.2) Limits of Overall Operating Loss

In cases where the angles and moduli of the reference inverse reflection coefficients and operating factors of the individual equipment units are known, the formulae in Section 5.1 are suitable for the determination of exact values of overall operating factors.

In production quality control and acceptance testing of

quantities of equipment it is often not practicable to determine the angles of the parameters mentioned above over the frequency range concerned. However, there are simple and speedy means of determining the moduli of individual equipment-unit inverse reflection coefficients and operating factors under reference conditions, and from such data it is possible to determine the limits of overall operating loss when the equipment units form part of a practical transmission system. These limits are of most use when impedance-matching conditions in the transmission system are relatively good.

The limits mentioned may be derived from the exact formulae of Section 5.1. Thus, considering the interacting case, the K -factor may be written as follows:

$$K = \frac{1}{1 - \frac{1}{A_S A_L}} \left[\left(1 - \frac{1}{A_S A_{R1}}\right) \left(1 - \frac{1}{A_{R2} A_L}\right) - \frac{1}{A_S A_L G^2} \right]$$

Although K is presented in terms of complex variables, the simplicity of the function permits the ready determination of

limits. Thus if k_{min} and k_{max} are, respectively, the minimum and maximum limits of $|K|$,

$$k_{max} = \frac{1}{1 - \frac{1}{a_S a_L}} \left[\left(1 + \frac{1}{a_S a_{R1}} \right) \left(1 + \frac{1}{a_{R2} a_L} \right) + \frac{1}{a_S a_L g^2} \right]$$

$$k_{min} = \frac{a_S a_L}{a_S a_L - 1} \left(\frac{a_S a_{R1} + 1}{a_S a_{R1}} \frac{a_{R2} a_L + 1}{a_{R2} a_L} + \frac{1}{a_S a_L g^2} \right)$$

which exists when

$$\theta_S + \theta_L = 2n_1\pi \quad (A_S A_L \text{ is real and positive})$$

$$\theta_S + \theta_{R1} = (2n_2 + 1)\pi \quad (A_S A_{R1} \text{ is real and negative})$$

$$\theta_{R2} + \theta_L = (2n_3 + 1)\pi \quad (A_{R2} A_L \text{ is real and negative})$$

$$\theta_S + \theta_L + 2\phi = (2n_4 + 1)\pi \quad (A_S A_L G^2 \text{ is real and negative})$$

These angular equations may be put into the form

$$\theta_S + \theta_L = 2n_1\pi$$

$$\theta_{R1} - \theta_L = [2(n_2 - n_1) + 1]\pi$$

$$\theta_{R2} + \theta_L = (2n_3 + 1)\pi$$

$$\phi = (n_4 - n_1)\pi + \frac{\pi}{2}$$

In this form it is clearly seen that the conditional angular equations for the existence of k_{max} may be satisfied.

Similarly, k_{min} may be shown to be

$$k_{min} = \frac{a_S a_L}{a_S a_L + 1} \left(\frac{a_S a_{R1} - 1}{a_S a_{R1}} \frac{a_{R2} a_L - 1}{a_{R2} a_L} - \frac{1}{a_S a_L g^2} \right)$$

This occurs when the right-hand factor is positive, which is the case in all normal applications, and

$$\theta_S + \theta_L = (2n_5 + 1)\pi$$

$$\theta_{R1} - \theta_L = [2(n_6 - n_5) - 1]\pi$$

$$\theta_{R2} + \theta_L = 2n_7\pi$$

$$\phi = (n_8 - n_5)\pi - \frac{\pi}{2}$$

The remainder of the formulae in Table 1 may be treated in a similar fashion, leading to the results set out in Table 2.

It will be seen that the expressions for k_{max} and k_{min} possess factors of the form $|X|/(|X| - 1)$ and $|X|/(|X| + 1)$, or the inverse of these. Defining $20 \log_{10} k$ as the variation loss, VL , in decibels, it follows that $VL_{max} = 20 \log_{10} k_{max}$ and $VL_{min} = 20 \log_{10} k_{min}$ will possess terms of the form $20 \log_{10} [|X|/(|X| - 1)]$ and $20 \log_{10} [|X|/(|X| + 1)]$. If $20 \log_{10} |X|$ is known, then with reference to Fig. 6, the value of $20 \log_{10} [|X|/(|X| - 1)]$ is obtained from the extreme upper curve and that of $20 \log_{10} [|X|/(|X| + 1)]$ from the extreme lower curve, so that VL_{max} and VL_{min} are readily computed.

Examples illustrating the application of the formulae given in Table 2 are presented in Section 6.

(6) APPLICATIONS OF OPERATING-LOSS LIMITS

The formulae applicable to the substitution of a quadripole for an ideal transducer or for another quadripole are in each case of the form

$$G'_{EE} = G_{EE} G K$$

where K is given in Table 1.

A method of obtaining the limits k_{min} and k_{max} of $k = |K|$ (as given in Table 2) has been demonstrated in Section 5.2. Further consideration shows that k may assume any value between these limits for substitution cases where the reference operating losses and return losses only are known. In certain cases additional information may be available which can be used to establish appropriate expressions for restricted limits (see Section 6.1).

In the general case, the expression for the overall operating factor quoted above may be used to establish formulae representing the frequency response of the related transmission system. Thus, introducing the variable frequency f ,

$$G'_{EE}(f) = G_{EE}(f) G(f) K(f)$$

or when considering moduli only,

$$g'_{EE}(f) = g_{EE}(f) g(f) k(f)$$

Introducing the limits of $k(f)$,

$$g_{EE}(f) g(f) k_{min}(f) \leq g'_{EE}(f) \leq g_{EE}(f) g(f) k_{max}(f)$$

For practical applications it is more convenient to adopt the logarithmic form of the last expression. Therefore, using $VL_{max} = 20 \log_{10} k_{max}$, $VL_{min} = 20 \log_{10} k_{min}$ (see Section 5.2) and $OL = 20 \log_{10} g$ (see Section 2),

$$OL_{EE}(f) + OL(f) + VL_{min}(f) \leq OL'_{EE}(f) \leq OL_{EE}(f) + OL(f) + VL_{max}(f)$$

where

$$VL_{min}(f) = 20 \log_{10} k_{min}(f) \leq 0, \text{ since } k_{min}(f) \leq 1$$

$$VL_{max}(f) = 20 \log_{10} k_{max}(f) \geq 0, \text{ since } k_{max}(f) \geq 1$$

This expression is completely general. For further treatment it is necessary to distinguish between replacements made with equipment individual to a single channel of a transmission system, and replacements made with equipment common to a group of channels. Treatment of the former case will be considered in detail; for the latter case results only are quoted, the treatment being similar.

For replacements made with equipment individual to a single channel, the channel reference frequency f_0 is normally common to all the quadripoles concerned, and equal to the line-up frequency, except where modulation and demodulation processes are involved. In these exceptional cases it will usually be found convenient to refer the frequencies to the input of the quadripole substituted.

Thus, applying the general expression at the line-up frequency f_0 ,

$$VL_{min}(f_0) \leq OL'_{EE}(f_0) - OL_{EE}(f_0) - OL(f_0) \leq VL_{max}(f_0)$$

The relevant frequency responses may be defined as

$$J(f) = OL(f_0) - OL(f)$$

$$J_{EE}(f) = OL_{EE}(f_0) - OL_{EE}(f)$$

$$J'_{EE}(f) = OL'_{EE}(f_0) - OL'_{EE}(f)$$

where, in accordance with general usage, a positive value of the relative operating-gain parameter $J(f)$ signifies that the operating gain at the frequency f is in excess of the operating gain at the reference frequency f_0 .

These definitions enable the following expression to be obtained:

$$J_{EE}(f) + J(f) - VL_{min}(f) + VL_{max}(f_0) \geq J'_{EE}(f) \geq J_{EE}(f) + J(f) - VL_{max}(f) + VL_{min}(f_0)$$

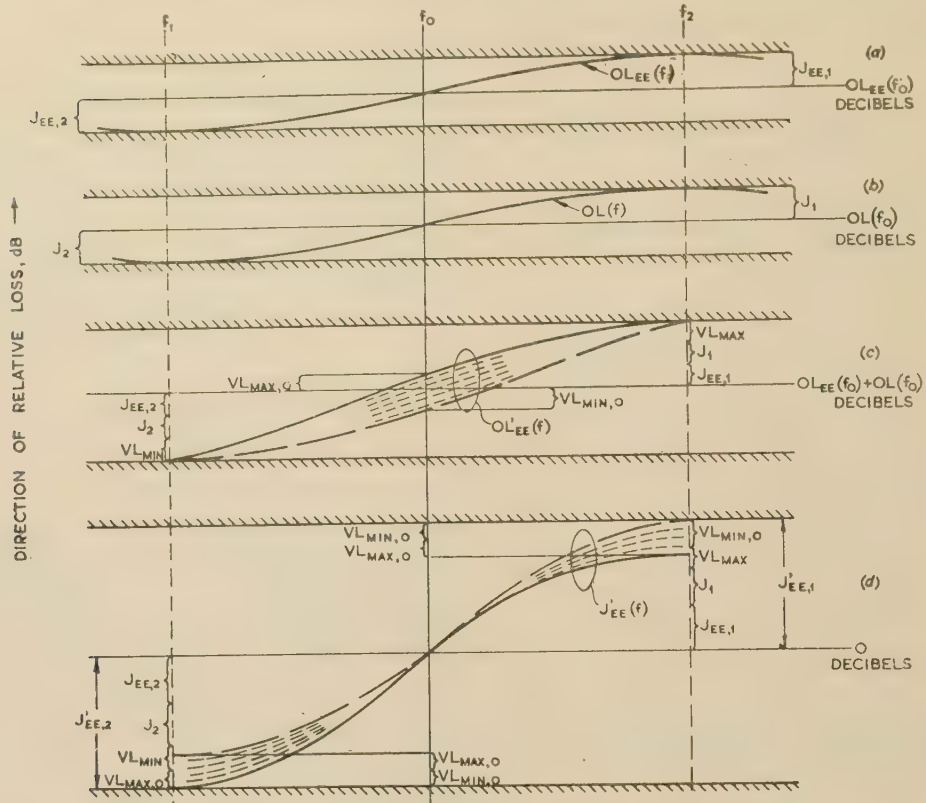


Fig. 7.—Frequency-response characteristics and limits.

- (a) Original channel operating-loss characteristic.
 (b) Operating-loss characteristic of quadripole to be used for replacement purposes.
 (c) New channel operating-loss characteristic, assuming VL_{min} to occur at f_1 and VL_{max} to occur at f_2 .
 (d) Channel frequency-response characteristic derived from (c).

This expression is useful for numerical calculations if $J_{EE}(f)$ and $J(f)$ are known. In many cases, however, the limits only of these parameters are known.

Thus

$$\begin{aligned} J_1 &\leq J(f) \leq J_2 \\ J_{EE,1} &\leq J_{EE}(f) \leq J_{EE,2} \\ J'_{EE,1} &\leq J'_{EE}(f) \leq J'_{EE,2} \end{aligned}$$

where J_1 , $J_{EE,1}$ and $J'_{EE,1}$ are negative gain constants, and J_2 , $J_{EE,2}$ and $J'_{EE,2}$ are positive gain constants. These constants are, in practice, termed the frequency-response limits.

Using these definitions the following is obtained:

$$\begin{aligned} J_{EE,2} + J_2 - VL_{min}(f) + VL_{max}(f_0) &\geq J'_{EE}(f) \\ &\geq J_{EE,1} + J_1 - VL_{max}(f) + VL_{min}(f_0) \end{aligned}$$

In the same manner that frequency-response limits are specified constant over certain frequency ranges, it is also usual to specify a constant minimum return-loss limit over a range of frequencies. By using such constant minimum return-loss limits and the constant minimum value of $OL(f)$, the minimum and maximum variation losses $VL_{min}(f)$ and $VL_{max}(f)$ become the constants VL_{min} and VL_{max} respectively. The reference frequency, f_0 , is normally situated near the centre of the channel frequency range, and the minimum return-loss limit at this frequency may therefore often be defined in excess of that at other frequencies. Where this is the case, the corresponding variation-loss limits may be represented by the symbols $VL_{min(0)}$ and $VL_{max(0)}$. Under these assumptions,

$$J'_{EE,1} = J_{EE,1} + J_1 - VL_{max} + VL_{min(0)}$$

and

$$J'_{EE,2} = J_{EE,2} + J_2 - VL_{min} + VL_{max(0)}$$

where

$$\begin{aligned} VL_{max} &> VL_{max(0)} \\ |VL_{min}| &> |VL_{min(0)}| \end{aligned}$$

From Fig. 7, which presents this result graphically, it will be seen that the actual frequency response $J'_{EE}(f)$ can touch one of the frequency-response limits only, and that the total clearance between the extreme turning points on the frequency-response characteristic and the frequency response-limits cannot become less than $VL_{max(0)} - VL_{min(0)}$.

The replacement of equipment units common to a group of channels may be treated in an identical manner, but since the reference frequency of the group equipment units will generally differ from that of the channel under consideration, the formulae for the frequency-response limits will differ from those obtained above. Thus, using the previous nomenclature, with the exception that $J(f)$ represents the frequency response of the group equipment unit concerned relative to the reference frequency of the group (which differs from the reference frequency of the channel), and defining J_1 and J_2 accordingly, it is possible to obtain

$$J'_{EE,1} = J_{EE,1} + J_1 - VL_{max} + VL_{min(0)} - J_2$$

and

$$J'_{EE,2} = J_{EE,2} + J_2 - VL_{min} + VL_{max(0)} - J_1$$

By comparing these results with those applicable to replacement of an equipment unit individual to a single channel, it is seen that $J'_{EE,1}$ and $J'_{EE,2}$ for the two cases differ by J_2 and J_1 , respectively. This is an inherent disadvantage of the particular method chosen for the specification of group equipment units, but this disadvantage can be avoided by specifying group equipment units in a manner somewhat similar to that adopted for equipment units individual to a single channel.

This alternative approach also permits the specification of smaller numerical values for J_1 and J_2 and may at the same time permit greater manufacturing tolerances. An example of its use is given in Section 7.1.

(6.1) Level-Adjustment Pad

If we consider equipment individual to a single channel, let it be assumed that a symmetrical level-adjustment pad is to be inserted at an intermediate impedance junction where $Z_{N1} = Z_{N2} = Z_N$, say. Under such conditions the straight-through connection may be regarded as an ideal transducer of unit ratio (see Sections 5 and 5.1.1). The replacement of the ideal transducer with the level-adjustment pad is illustrated in Fig. 4.

To simplify the example, it will be assumed that the nominal impedances are of zero angle, and that the level-adjustment pad is of good quality, so that

$$Z_N = Z_{R1} = Z_{R2}$$

and consequently $|A_{R1}| = |A_{R2}| = \infty$.

Examination of Table 2 for the case where the reference parameters become the image reference parameters shows that, for this example, the conditional angular equations cannot be satisfied, since the angle of G_i is here zero, and not $\pi/2$ as required. The limits k_{min} and k_{max} of k , as tabulated in Table 2, are therefore not applicable. In order to establish the restricted limits of k for $\phi = n\pi$, reference is made to Table 1, which gives the following formula:

$$K = \frac{A_S A_L}{A_S A_L - 1} \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2}$$

This may be rearranged as follows:

$$K = 1 + \frac{1 - \frac{1}{G_i^2}}{A_S A_L - 1}$$

Since G_i^2 must be real and positive it is readily seen that restricted limits of k are obtained as follows:

$$k_{max} = \frac{a_S a_L}{a_S a_L - 1} \frac{a_S a_L g_i^2 - 1}{a_S a_L g_i^2}, \text{ for } \theta_S + \theta_L = 2n\pi$$

and

$$k_{min} = \frac{a_S a_L}{a_S a_L + 1} \frac{a_S a_L g_i^2 + 1}{a_S a_L g_i^2}, \text{ for } \theta_S + \theta_L = (2n + 1)\pi$$

If we assume the following data:

Image reference operating loss, $OL_i = 20 \log_{10} g_i = 3$ dB, $RL_S(f) \geq 10$ dB, $RL_S(f_0) \geq 16$ dB, $RL_L(f) \geq 20$ dB and $RL_L(f_0) \geq 26$ dB. Then, using the lower limit of return loss in each case:

At f_0 : $20 \log_{10} (a_S a_L) = RL_S(f_0) + RL_L(f_0) = 42$ dB, $20 \log_{10} (a_S a_L g_i^2) = RL_S(f_0) + RL_L(f_0) + 2OL_i = 48$ dB, and from large-scale graphs similar to Fig. 6,

$$VL_{max(0)} = 0.07 - 0.03 = 0.04 \text{ dB}$$

$$VL_{min(0)} = -0.07 + 0.03 = -0.04 \text{ dB}$$

similarly, at f :

$$VL_{max} = 0.14 \text{ dB}$$

and

$$VL_{min} = -0.14 \text{ dB}$$

The loss $OL'_{EE}(f_0)$ at the line-up frequency thus becomes

$$OL'_{EE}(f_0) + 2.96 \leq OL'_{EE}(f_0) \leq OL_{EE}(f_0) + 3.04 \text{ decibels}$$

and the new frequency-response limits are

$$J'_{EE,1} = J_{EE,1} + 0 - 0.14 - 0.04 = J_{EE,1} - 0.18 \text{ decibels}$$

$$J'_{EE,2} = J_{EE,2} + 0 + 0.14 + 0.04 = J_{EE,2} + 0.18 \text{ decibels}$$

It is of interest to note that, if the angle of the operating factor had not been known, the limiting values k_{max} and k_{min} as presented in Table 2 would have been applicable. In this case the following results are obtained:

$$OL_{EE}(f_0) + 2.9 \leq OL'_{EE}(f_0) \leq OL_{EE}(f_0) + 3.1 \text{ decibels}$$

$$J'_{EE,1} = J_{EE,1} + 0 - 0.41 - 0.1 = J_{EE,1} - 0.51 \text{ decibels}$$

$$J'_{EE,2} = J_{EE,2} + 0 + 0.41 + 0.1 = J_{EE,2} + 0.51 \text{ decibels}$$

These results have deteriorated significantly as compared with those obtained for the level-adjustment pad.

(7) SYNTHESIS OF SIMPLE TRANSMISSION SYSTEMS

In the previous Sections the emphasis has been on replacement or substitution of quadripoles. It will be realized that complete transmission systems may be synthesized by performing a number of such substitutions. The method for simple cases may be developed as follows. Consider the transmission system composed of ideal transducers, as shown in Fig. 8(a). The three ideal transducers effect image matching at all impedance junctions, and the overall operating factor is consequently unity. As will be observed, it is convenient to designate the nominal impedances as Z_{N1} , Z_{N3} , Z_{N5} , etc., rather than Z_{N1} , Z_{N2} ,

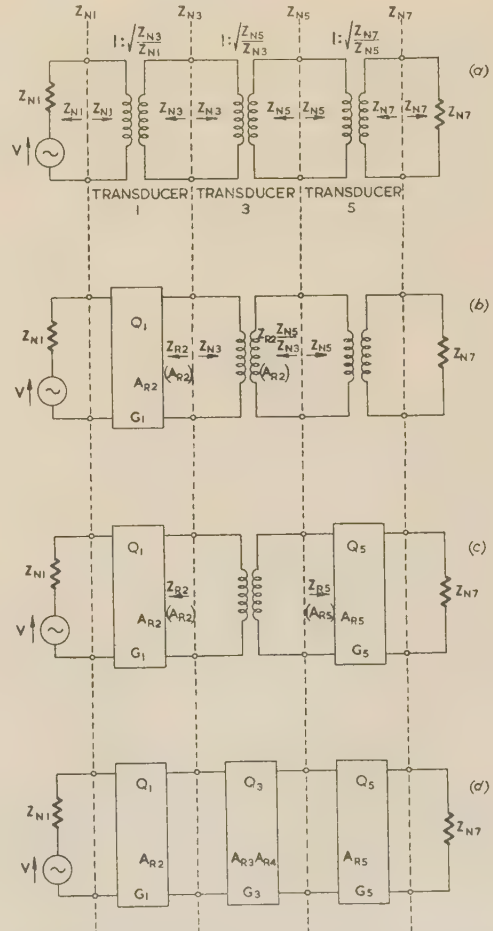


Fig. 8.—Stages in the synthesis of a simple transmission system.

Z_{N3} , etc., as before, since this permits the reference parameters of the quadripole Q_n operating between Z_{Nn} and $Z_{N(n+2)}$ to be simply expressed as G_n , Z_{Rn} and $Z_{R(n+1)}$, where n is an odd integer. Assuming $n = 1$, the ideal transducer 1 is now replaced by quadripole Q_1 . The source and load impedances for this quadripole are Z_{N1} and Z_{N3} , respectively. To permit calculation G_1 and Z_{R2} (or A_{R2}) must be known. Considering Fig. 8(b), the overall operating factor at this stage is obviously G_1 , since Q_1 is operating between its appropriate nominal impedances.

If we assume that $n = 5$, the ideal transducer 5 is replaced by quadripole Q_5 . The nominal source and load impedances for Q_5 are Z_{N5} and Z_{N7} , respectively, and calculation may proceed if G_5 and Z_{R5} (or A_{R5}) are known. From Fig. 8(c), the quadripole Q_5 will operate between the source impedance $Z_{R2}Z_{N5}/Z_{N3}$ and the load impedance Z_{N7} . The inverse reflection coefficient $Z_{R2}Z_{N5}/Z_{N3}$ against Z_{N5} is written as follows:

$$A = \frac{\frac{Z_{R2}Z_{N5}}{Z_{N3}} + 1}{\frac{Z_{R2}Z_{N5}}{Z_{N3}} - 1} = \frac{\frac{Z_{R2}}{Z_{N3}} + 1}{\frac{Z_{R2}}{Z_{N3}} - 1} = A_{R2}$$

Using this result, and the relevant formula from Table 1, the overall operating factor at this stage is found to be

$$G_{EE,S} = G_1 G_5 \frac{A_{R2} A_{R5} - 1}{A_{R2} A_{R5}}$$

Finally, for $n = 3$ the ideal transducer 3 is replaced by the quadripole Q_3 . The nominal source and load impedances for Q_3 are Z_{N3} and Z_{N5} respectively. For calculation, the reference parameters G_3 , Z_{R3} (or A_{R3}) and Z_{R4} (or A_{R4}) must be known. Considering Fig. 8(d), the overall operating factor $G'_{EE,S}$ of the synthesized transmission system may, by use of the appropriate formulae from Table 1 and the nomenclature adopted in this Section, be expressed as follows:

$$G'_{EE,S} = G_1 G_5 \frac{A_{R2} A_{R5} - 1}{A_{R2} A_{R5}} G_3 \frac{A_{R2} A_{R5}}{A_{R2} A_{R5} - 1} X \frac{X A_{R2} A_{R5} G_3^2 - 1}{X A_{R2} A_{R5} G_3^2} \\ = G_1 G_3 G_5 X \frac{X A_{R2} A_{R5} G_3^2 - 1}{X A_{R2} A_{R5} G_3^2}$$

where
$$X = \frac{A_{R2} A_{R3} - 1}{A_{R2} A_{R3}} \frac{A_{R4} A_{R5} - 1}{A_{R4} A_{R5}}$$

From this general formula, applicable to a transmission system consisting of three quadripoles connected in tandem, the formulae for two special cases may be derived.

Special Case 1.

Quadripole Q_3 non-interacting.

$$G'_{EE,S} = G_1 G_3 G_5 X$$

Special Case 2.

Synthesis of a transmission system consisting of two quadripoles only.

In this case $|A_{R3}| = |A_{R4}| = \infty$, $G_3 = 1$ and $Z_{N3} = Z_{N5}$.

Then
$$G_{EE,S} = G_1 G_5 \frac{A_{R2} A_{R5} - 1}{A_{R2} A_{R5}}$$

Further, using these formulae, transmission systems consisting of more than three quadripoles may be synthesized, provided that the quadripoles are arranged in such a manner that

(a) There are not more than two interacting quadripoles in tandem at each end of the system.

(b) The remaining quadripoles in the system are all of the non-

interacting type, or, if interacting quadripoles are included, the quadripoles must be so arranged that not more than one interacting quadripole separates two non-interacting quadripoles.

To obtain the formula for the overall operating factor, the transmission system having an overall operating factor of unity is formed from the appropriate number of ideal transducers. Each of the ideal transducers is then replaced in turn by a quadripole in such a fashion that all non-interacting quadripoles are inserted before interacting quadripoles. In this process, quadripoles possessing a high reference operating loss may be considered non-interacting without introduction of appreciable error.

The formulae for the overall operating factor of transmission systems synthesized from two and three quadripoles are presented in Table 3, together with the limits of k for these cases.

Table 3

SYNTHESIS OF TRANSMISSION SYSTEMS CONSISTING OF TWO AND THREE QUADRIPOLES AS SHOWN IN FIG. 8

Definitions of auxiliary parameters used in this Table	$X = \frac{A_{R2} A_{R3} - 1}{A_{R2} A_{R3}} \frac{A_{R4} A_{R5} - 1}{A_{R4} A_{R5}}$ $x_{max} = \frac{a_{R2} a_{R3} + 1}{a_{R2} a_{R3}} \frac{a_{R4} a_{R5} + 1}{a_{R4} a_{R5}}$ <p>for $\begin{cases} \theta_{R2} + \theta_{R3} = (2n_1 + 1)\pi \\ \theta_{R4} + \theta_{R5} = (2n_2 + 1)\pi \end{cases}$ (1)</p> $x_{min} = \frac{a_{R2} a_{R3} - 1}{a_{R2} a_{R3}} \frac{a_{R4} a_{R5} - 1}{a_{R4} a_{R5}}$ <p>for $\begin{cases} \theta_{R2} + \theta_{R3} = 2n_1\pi \\ \theta_{R4} + \theta_{R5} = 2n_2\pi \end{cases}$ (2)</p>
Three Quadripoles Middle quadripole interacting	$G'_{EE,S} = G_1 G_3 G_5 K,$ <p>where $K = X \frac{X A_{R2} A_{R5} G_3^2 - 1}{X A_{R2} A_{R5} G_3^2}$</p> <p>Limits of $k = K$:</p> $k_{max} = x_{max} \frac{x_{max} a_{R2} a_{R5} g_3^2 + 1}{x_{max} a_{R2} a_{R5} g_3^2}$ <p>if eqn. (1) is met, and</p> $\theta_{R2} + \theta_{R5} + 2\phi_3 = (2n_3 + 1)\pi$ $k_{min} = x_{min} \frac{x_{min} a_{R2} a_{R5} g_3^2 - 1}{x_{min} a_{R2} a_{R5} g_3^2}$ <p>if eqn. (2) is met, and</p> $\theta_{R2} + \theta_{R5} + 2\phi_3 = 2n_3\pi$
Three Quadripoles Middle quadripole non-interacting	$G_{EE,S} = G_1 G_3 G_5 K, \text{ where } K = X$ <p>Limits of $k = K$:</p> $k_{max} = x_{max}, \text{ if eqn. (1) is met}$ $k_{min} = x_{min}, \text{ if eqn. (2) is met}$
Two Quadripoles	$G_{EE,S} = G_1 G_5 K,$ <p>where $K = \frac{A_{R2} A_{R5} - 1}{A_{R2} A_{R5}}$</p> <p>Limits</p> $k_{max} = \frac{a_{R2} a_{R5} + 1}{a_{R2} a_{R5}}$ <p>if $\theta_{R2} + \theta_{R5} = (2n_1 + 1)\pi$</p> $k_{min} = \frac{a_{R2} a_{R5} - 1}{a_{R2} a_{R5}}$ <p>if $\theta_{R2} + \theta_{R5} = 2n_1\pi$</p>

(7.1) Synthesis of a Cable Carrier Telephone Terminal

In Section 6 it was suggested that the generally adopted method of specifying frequency responses of equipment common to a group of channels could be improved. The revised method suggested by the authors is demonstrated in this practical example.

It is assumed that the carrier telephone terminal consists of three equipment units, as shown in Fig. 9. The channel modem*

Frequency-Response Limits.

$$OL_3(f_{0,n}) - 0.5 \leq OL_3(f_n) \leq OL_3(f_{0,n})$$

$$+ \begin{cases} 4.0 \text{ (dB)}, 300 \leq f_n < 400 \text{ c/s} \\ 1.5 \text{ (dB)}, 400 \leq f_n < 600 \text{ c/s} \\ 0.5 \text{ (dB)}, 600 \leq f_n \leq 2400 \text{ c/s} \\ 1.5 \text{ (dB)}, 2400 < f_n \leq 3000 \text{ c/s} \\ 4.0 \text{ (dB)}, 3000 < f_n \leq 3400 \text{ c/s} \end{cases}$$

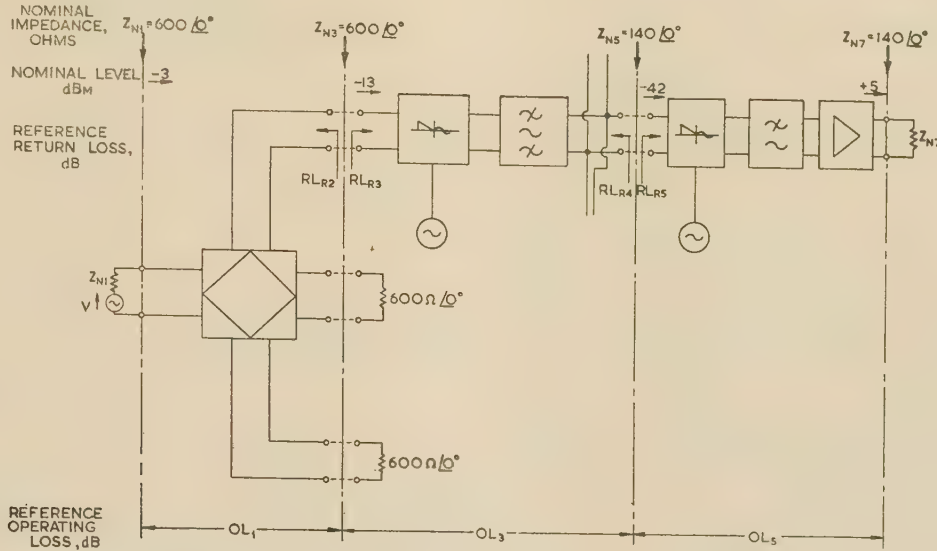


Fig. 9.—Synthesis of a cable carrier telephone terminal.

and group translating equipments will generally incorporate rectifier modulators, which differ in behaviour from linear passive quadripoles. However, if such a modulator is suitably designed^{3,4,5} and associated with a non-interacting or high-loss quadripole, the equipment unit so obtained can, with negligible error, be regarded as a linear non-interacting quadripole. This assumption is made in this example, where, for simplicity, the transmitting path only is considered.

Considering, therefore, the n th channel of the group, the required performance characteristics of the equipment units of Fig. 9 may be specified as follows:

(a) Two-Wire/Four-Wire Terminating Equipment.

Return and Operating Losses.

$$RL_{R2}(f_{0,n}) \geq 26 \text{ dB, and } RL_{R2}(f_n) \geq 20 \text{ dB} \\ 10 - 0.25 \leq OL_1(f_{0,n}) \leq 10 + 0.25 \text{ dB}$$

Frequency-Response Limits.

$$OL_1(f_{0,n}) - 0.1 \leq OL_1(f_n) \leq OL_1(f_{0,n}) \\ + \begin{cases} 0.2 \text{ (dB)}, 300 \leq f_n < 600 \text{ c/s} \\ 0.1 \text{ (dB)}, 600 \leq f_n \leq 3000 \text{ c/s} \\ 0.2 \text{ (dB)}, 3000 < f_n \leq 3400 \text{ c/s} \end{cases}$$

(b) Channel Modem Equipment.

Return and Operating Losses.

$$RL_{R3}(f_{0,n}) \geq 26 \text{ dB, and } RL_{R3}(f_n) \geq 20 \text{ dB} \\ RL_{R4}(f_{0,n}) \geq 16 \text{ dB, and } RL_{R4}(f_n) \geq 10 \text{ dB} \\ 29 - 0.25 \leq OL_3(f_{0,n}) \leq 29 + 0.25 \text{ dB}$$

(c) Group Translating Equipment.

Return and Operating Losses.

$$RL_{R5}(f_G) \geq 20 \text{ dB, } RL_{R5}(f_{0,n}) \geq 20 \text{ dB, and } RL_{R5}(f_n) \geq 20 \text{ dB} \\ -47 - 0.25 \leq OL_5(f_G) \leq -47 + 0.25 \text{ dB}$$

Frequency-Response Limits.

$$OL_5(f_G) - 0.25 \leq OL_5(f_{0,n}) \leq OL_5(f_G) + 0.25 \text{ dB} \\ OL_5(f_{0,n}) - 0.1 \leq OL_5(f_n) \leq OL_5(f_{0,n}) + 0.1 \text{ dB}$$

The frequency-response limits given under (a), (b) and (c) are presented in Fig. 10.

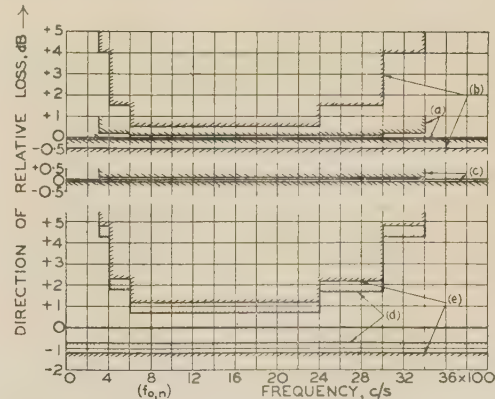


Fig. 10.—Frequency-response limits relating to Fig. 9.

- (a) Two-wire/four-wire termination.
- (b) Channel modem equipment.
- (c) Group translating equipment.
- (d) Summation of (a), (b) and (c).
- (e) Complete terminal.

* The term 'modem' is a convenient contraction of 'modulator-demodulator'.

The conditional angular equations for the existence of the limits of k , as given in Table 3, may be satisfied. Thus, using the lower return-loss limits,

$$VL_{max}(f_{0n}) = 0.02 + 0.14 = 0.16 \text{ dB}$$

$$VL_{min}(f_{0n}) = -0.02 - 0.14 = -0.16 \text{ dB}$$

$$VL_{max}(f_n) = 0.09 + 0.26 = 0.35 \text{ dB}$$

$$VL_{min}(f_n) = -0.09 - 0.28 = -0.37 \text{ dB}$$

Therefore, at the channel reference frequency f_{0n} ,

$$\begin{aligned} (10 - 0.25) + (29 - 0.25) + (-47 - 0.25 - 0.25) - 0.16 \\ \leq OL'_{EE,S}(f_{0n}) \leq (10 + 0.25) + (29 + 0.25) \\ + (-47 + 0.25 + 0.25) + 0.16 \text{ dB, or} \\ -9.16 \leq OL'_{EE,S}(f_{0n}) \leq -6.84 \text{ (dB)} \end{aligned}$$

If an input signal level of -3 dBm is assumed at the hybrid line, the level at the output of the transmitting amplifier will then fall between the limits $+3.84 \text{ dBm}$ and $+6.16 \text{ dBm}$ and will have a nominal value of $+5 \text{ dBm}$. This range of 2.3 dB allows for a tolerance of $\pm 0.25 \text{ dB}$ on $OL_5(f_G)$, and it will therefore be clear that the maximum difference between any two channels cannot exceed 1.8 dB .

The overall channel frequency-response limits shown in Fig. 10 are obtained by widening the direct sum of the upper limits of the three equipment units by $VL_{min}(f_{0n}) - VL_{max}(f_n) = -0.51 \text{ dB}$, and the direct sum of the lower limits by $VL_{max}(f_{0n}) - VL_{min}(f_n) = 0.53 \text{ dB}$.

The method of specification suggested here has advantages compared with other methods in current use, since it permits improved overall frequency-response limits to be obtained without demanding improved performance from the individual equipment units.

It will be noticed in this example that under reference conditions the operating loss from 'hybrid line' to 'hybrid out', and the 'hybrid out' impedance have been considered invariable with changes in the impedance terminating 'hybrid in'. These assumptions are correct for ideal hybrid transformers, and practical measurements have confirmed the validity of the theoretical proof. A similar invariance applies to the operating loss from 'hybrid in' to 'hybrid line', and to the 'hybrid in' impedance.

(8) INPUT INVERSE REFLECTION COEFFICIENT OF AN INTERACTING QUADRIPOLE UNDER NON-REFERENCE CONDITIONS

The input impedance Z_{in} of an interacting quadripole Q , terminated at the far end with an impedance Z_L differing from

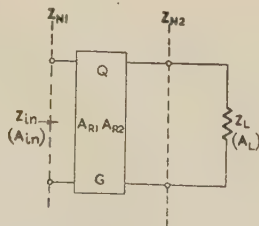


Fig. 11.—Input impedance of an interacting quadripole under non-reference conditions.

the nominal impedance Z_{N2} , has been obtained in Section 15.4 in terms of the parameters illustrated in Fig. 11. Using this result,

$$\frac{Z_{in} - 1}{Z_{N1} + 1} = \frac{Z_{R1} - 1}{Z_{N1} + 1} + \frac{\left(\frac{Z_{R2} + 1}{Z_{N2}}\right)\left(\frac{Z_L - 1}{Z_{N2}}\right)}{2\left(\frac{Z_{R2}}{Z_{N2}} + \frac{Z_L}{Z_{N2}}\right)G^2}$$

By introducing inverse reflection coefficients, this reduces to

$$A_{in} = A_{R1} \frac{\frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}}{\frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L} + 1} = A_{R1} \frac{X}{X + 1}$$

where

$$X = \frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}$$

Let it now be assumed that $X = x e^{j\alpha}$, so that

$$\frac{X}{X + 1} = \frac{x e^{j\alpha}}{x e^{j\alpha} + 1}$$

Then multiplying by the unit ratio $\frac{e^{j\pi}}{e^{j\pi}}$

$$\frac{X}{X + 1} = \frac{x e^{j(\alpha + \pi)}}{x e^{j(\alpha + \pi)} - 1} = \frac{X'}{X' - 1}$$

where $X' = x e^{j(\alpha + \pi)}$.

Thus, by working in terms of $20 \log_{10} x$ decibels and $(\alpha + 180^\circ)$, the curves of Fig. 6 may be used to evaluate $20 \log_{10} |X/(X + 1)|$ decibels and the angle of $X/(X + 1)$.

The results obtained here, and those for the special case of $Z_{R2} = Z_{N2}$ (or $|A_{R2}| = \infty$) are given in Table 4. Although in practice Z_{R2} generally differs from Z_{N2} , it will be appreciated that the results obtained from the general formula and those obtained from the special-case formula do not differ appreciably if $|A_{R2} A_L| \gg 1$.

When the reference parameters may be identified as the image reference parameters, the formula simply reduces to

$$A_{in} = A_L G_i^2$$

In a manner similar to that used for the determination of the limits of k , the minimum limits of a_{in} , corresponding, respectively, to the existence of k_{min} and k_{max} , are obtained, and appear in Table 4.

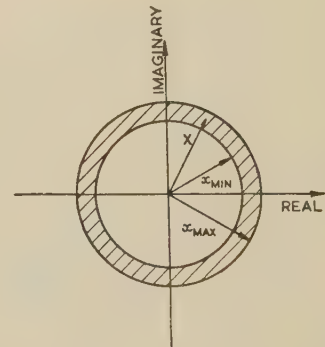


Fig. 12.—Locus area of the complex quantity.

$$\begin{aligned} X &= \frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L} \\ x_{min} &= \frac{a_L G^2}{a_{R1}} \frac{a_{R2} a_L - 1}{a_{R2} a_L}; \quad x_{max} = \frac{a_L G^2}{a_{R1}} \frac{a_{R2} a_L + 1}{a_{R2} a_L} \end{aligned}$$

In most practical cases, $x_{min} \approx |X| \approx x_{max} = \frac{a_L G^2}{a_{R1}}$

Table 4

INPUT INVERSE REFLECTION COEFFICIENTS (Fig. 11)

(Conditional angular equations omitted)

	General case	Special case $Z_{R2} = Z_{N2}$, or $ A_{R2} = \infty$
Definitions of auxiliary parameters used in this Table ..	$X = \frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}$ $x_{min} = \frac{a_L g^2}{a_{R1}} \frac{a_{R2} a_L - 1}{a_{R2} a_L}$ $x_{max} = \frac{a_L g^2}{a_{R1}} \frac{a_{R2} a_L + 1}{a_{R2} a_L}$	$X = \frac{A_L G^2}{A_{R1}}$ $x = \frac{a_L g^2}{a_{R1}}$
Exact formula	$A_{in} = A_{R1} \frac{X}{X + 1}$	$A_{in} = A_{R1} \frac{X}{X + 1}$
Minimum limit (corresponds to the existence of k_{min}) ..	$a_{in,min} = a_{R1} \frac{x_{min}}{x_{min} + 1}$	$a_{in,min} = a_{R1} \frac{x}{x + 1}$ (See Fig. 13)
Restricted minimum limit (corresponds to the existence of k_{max})	$a_{in,res.min} = a_{R1} \frac{x_{max}}{x_{max} + 1}$	$a_{in,res.min} = a_{in,min}$
Maximum limit	$a_{in,max} = a_{R1} \left \frac{x_{max}}{x_{max} - 1} \right $ <p>if $x_{max} \leq 1$</p> $a_{in,max} = \infty \text{ may occur}$ <p>if $x_{min} \leq 1 \leq x_{max}$</p> $a_{in,max} = a_{R1} \frac{x_{min}}{x_{min} - 1}$ <p>if $1 \leq x_{min}$</p>	$a_{in,max} = a_{R1} \left \frac{x}{x - 1} \right $ (See Fig. 13)

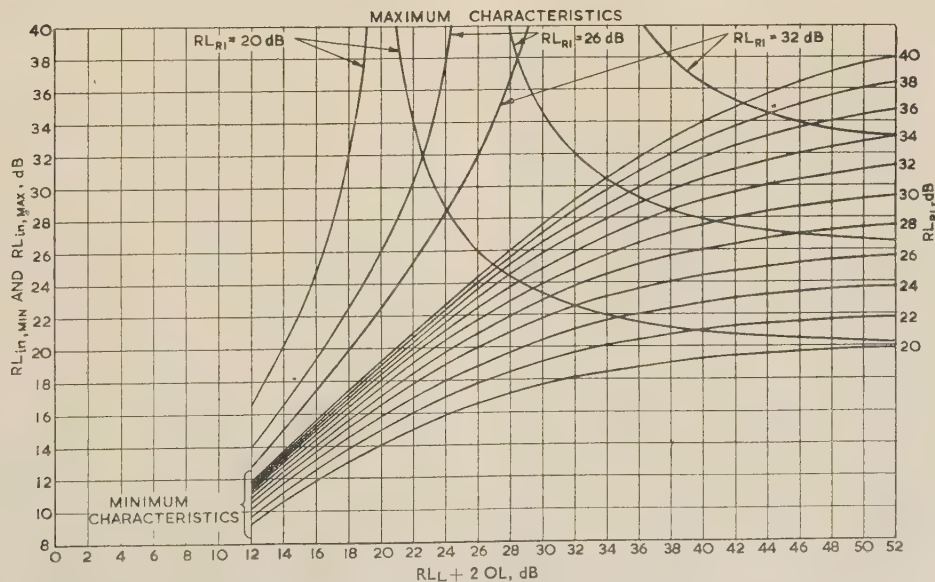


Fig. 13.—Input-return-loss limits of an interacting quadripole under non-reference conditions.

To determine the maximum limit of a_{in} , reference is made to Fig. 12, in which the locus area of the complex number

$$X = \frac{A_L G^2}{A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L}$$

is presented. Consideration of X as a negative real number in the region of $(-1 + j0)$ yields the formulae shown in Table 4.

Numerical calculations of limits of $RL_{in} = 20 \log_{10} a_{in}$ decibels in cases where it is necessary to use the general formulae, may conveniently be carried out by means of the curves of Fig. 6.

In cases where $|A_{R2} A_L| \gg 1$, $RL_{in,min}$ and $RL_{in,max}$ may be obtained directly from the family of curves in Fig. 13.

The formulae developed may be rearranged to solve other problems, and it is found that any chosen parameter may be expressed explicitly in terms of the remaining parameters in such a form that numerical calculations can be carried out using the curves of Fig. 6. Thus the formulae developed may be used to solve the important practical problem of specifying the reference-loss parameters of a quadripole in such a manner that the input return loss exceeds a given minimum value when the far-end termination differs from the far-end nominal impedance (see Section 10.1.2).

(8.1) Attenuator

Using the formula $A_{in} = A_L G_i^2$, which applies when the reference parameters may be identified as the image reference parameters, it is seen that

$$RL_{in} = RL_L + 2 OL_i \text{ decibels}$$

where RL_{in} = Input return loss of the attenuator.

RL_L = Return loss of the load impedance against the far-end nominal impedance.

OL_i = Image reference operating loss of the attenuator.

When the load impedance is either a pure inductance or a pure capacitance, $RL_L = 0$ and $RL_{in} = 2 OL_i$. This result has been found of use when checking the accuracy of return-loss measuring sets.

(9) SYNTHESIS OF MORE COMPLICATED TRANSMISSION SYSTEMS

If no restrictions are placed upon the arrangement of quadripoles forming a transmission system, so that chains of tandem-connected interacting quadripoles may be encountered, an explicit mathematical expression for the overall operating factor becomes extremely complicated.

However, for calculation purposes it is possible to employ a step-by-step numerical process, evaluating the overall operating loss progressively as each quadripole is added to the system. The method adopted is that used for the synthesis of simple transmission systems, with the extension that, when no further progress can be made by that method, it becomes necessary to evaluate RL_{in} (or $RL_{in,min}$ and $RL_{in,res,min}$ in the case of limit calculations) using the formulae given in Table 4. In this manner all ideal transducers can be replaced by quadripoles and the final overall operating loss can be obtained.

In the case of limit calculations, if data in excess of the reference operating and return losses are known for a quadripole, it may become necessary to determine the appropriate restricted limits of operating loss in a manner similar to that shown in Section 6.1 for the level-adjustment pad. It should be borne in mind that, as the number of quadripoles in a transmission system increases, the likelihood that they will simultaneously function under limiting operating-loss conditions decreases, and the calculation of operating-loss limits therefore becomes of less practical use.

(10) REFLECTIONS AT A JUNCTION OF DISSIMILAR IMPEDANCES

Considering a junction in a transmission system at which the nominal impedance is Z_N and the dissimilar junction impedances are Z_l and Z_r , the inverse reflection coefficients A_J , A_l and A_r become

$$A_J = \frac{Z_r + Z_l}{Z_r - Z_l}, \quad A_l = \frac{Z_l + Z_N}{Z_l - Z_N}, \quad A_r = \frac{Z_r + Z_N}{Z_r - Z_N}$$

If Z_l and Z_r are derived from the last two expressions and inserted into the first, it is seen that

$$A_J = \frac{A_l A_r - 1}{A_l - A_r} = A_r \frac{A_l A_r - 1}{A_l A_r} \frac{A_l}{A_l - 1}$$

From this expression the limits of a_J are found to be

$$a_{J,max} = a_r \frac{a_r a_l + 1}{a_r a_l} \left| \frac{\frac{a_l}{a_r}}{\frac{a_l}{a_r} - 1} \right|$$

$$\text{for } \theta_l = (n_1 + n_2)\pi + \frac{\pi}{2}$$

$$\text{and } \theta_r = (n_1 - n_2)\pi + \frac{\pi}{2}$$

$$\text{and } a_{J,min} = a_r \frac{a_r a_l - 1}{a_r a_l} \frac{\frac{a_l}{a_r}}{\frac{a_l}{a_r} + 1}$$

$$\text{for } \theta_l = (n_3 + n_4)\pi + \frac{\pi}{2}$$

$$\theta_r = (n_3 - n_4)\pi - \frac{\pi}{2}$$

A_J and RL_J may be calculated in the same manner as the K 's in the previous Sections, while $RL_{J,max}$ and $RL_{J,min}$ may either be calculated in the same manner as k_{max} and k_{min} , or obtained from the curves given in Fig. 14.

(10.1) Open-Wire Entrance Arrangement

Where open-wire carrier telephone systems are used it is necessary to provide lead-in arrangements to intermediate repeaters and terminal equipment. If both 3-channel and 12-channel frequency groups are propagated over the open-wire bearer, line filter groups are included in the lead-in arrangements to effect segregation of these frequency groups.

(10.1.1) Practical Example of a Line Filter Group.

The reference-loss parameters of a line filter group were measured in the transmission band of the high-pass filter section. The filter—an experimental model—was to be mounted on a terminal pole of the open-wire route, and the output of the high-pass section extended to the carrier station by means of an unloaded pair in the entrance cable. The reference-return-loss/frequency characteristic of this cable was measured, and using the results of Table 4 a prediction was made of the maximum and minimum return-loss/frequency limits possible at the line terminals of the filter group, assuming the unloaded entrance cable pair to be connected to the output of the high-pass section.

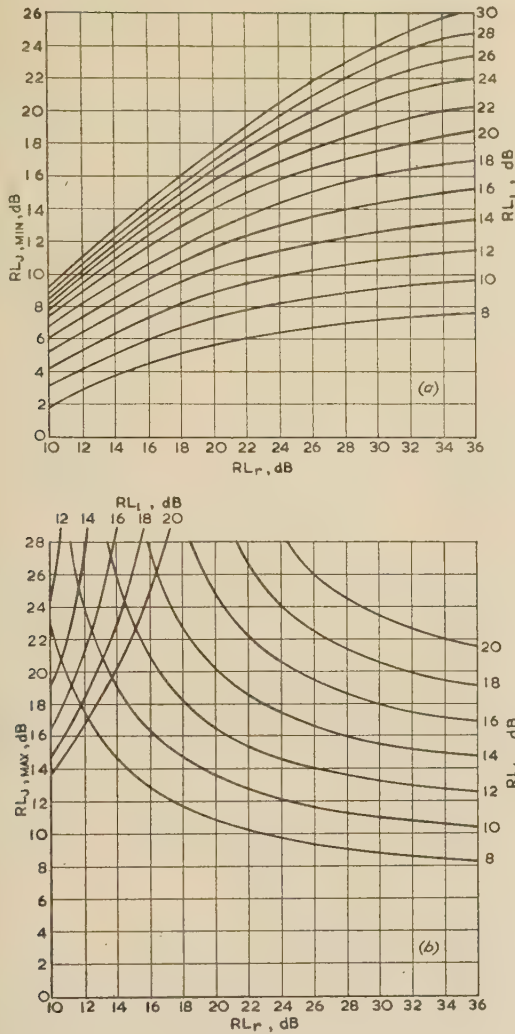


Fig. 14.—Return-loss limits at an impedance junction.

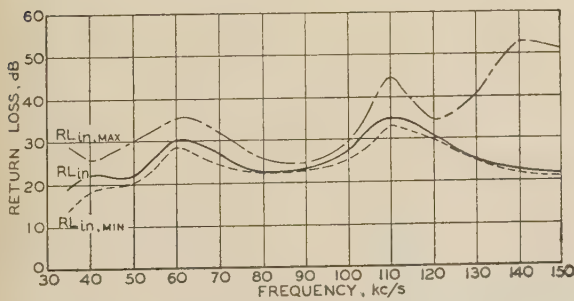


Fig. 15.—Measured input-return-loss characteristic of a line filter group, together with calculated upper and lower return-loss limits.

— RL_{in}
 --- $RL_{in, \min}$
 $RL_{in, \max}$

These maximum and minimum return-loss/frequency limits are shown in Fig. 15, together with the actual return-loss/frequency characteristic measured at the line terminals of the filter group under the conditions cited.

(10.1.2) Specification of Lower Return-Loss/Frequency Limits.

The reflection conditions existing at the impedance junction formed by connection of the open-wire pair to the line terminals

of the filter group may be assessed quantitatively by means of the formulae developed in Section 10.

Crosstalk considerations impose a minimum limit on the return loss required at this impedance junction, and, as mentioned in Section 6, it is convenient to quote this minimum limit as a constant return loss over the frequency band concerned (or, if necessary, as a number of constant return losses corresponding to different frequency intervals in this band). In this example a single minimum return-loss limit $RL_{J, \min}$ will be assumed.

If, therefore, at any frequency the return-loss characteristic falls below the minimum return-loss limit $RL_{J, \min}$ chosen—and this is to be avoided—a rearrangement of the results of Sections 8 and 10 permits calculation of the reference-loss parameters of another line filter group, which, when terminated with the given entrance cable, would provide a return loss at the junction of open-wire pair (of known reference return-loss/frequency characteristic) and line terminals of the filter group equal to or in excess of the prescribed limit.

The process is described briefly as follows:

(i) Considering the known data, it is first convenient to specify, for the open-wire and unloaded entrance cable pairs, constant minimum reference return-loss limits above which the respective measured return-loss characteristics must lie. Let these minimum limits be $RL_{ow} = 20 \log_{10} a_{ow}$ decibels and $RL_{EC} = 20 \log_{10} a_{EC}$ decibels, and let $RL_{J, \min} = 20 \log_{10} a_{J, \min}$ decibels.

(ii) Using the nomenclature above, the expression in Section 10 for $a_{J, \min}$ may be rearranged to give

$$a_r = a_{J, \min} \frac{a_{J, \min} a_{ow} + 1}{a_{J, \min} a_{ow}} \frac{\frac{a_{ow}}{a_{J, \min}}}{\frac{a_{ow}}{a_{J, \min}} - 1}$$

From the data given in (i), $RL_r = 20 \log_{10} a_r$ may readily be calculated using the extreme upper and lower curves of Fig. 6, or from Tables.

(iii) RL_r is the minimum permissible return loss at the line terminals of the filter group when terminated with the entrance cable pair. Thus a_r is the same as $a_{in, \min}$ in Table 4, and by rearrangement of the expression for $a_{in, \min}$ the following is obtained:

$$a_{R1} = a_r \frac{\frac{a_{EC}^2}{a_r} \frac{a_{R2} a_{EC} - 1}{a_{R2} a_{EC}}}{\frac{a_{EC}^2}{a_r} \frac{a_{R2} a_{EC} - 1}{a_{R2} a_{EC}} - 1}$$

which approximates to

$$a_{R1} = a_r \frac{\frac{a_{EC}^2}{a_r}}{\frac{a_{EC}^2}{a_r} - 1}$$

for the condition $a_{R2} a_{EC} \gg 1$, which is often encountered in practice.

If now $OL = 20 \log_{10} g$ decibels is allocated a constant value based on the minimum reference operating loss anticipated in the transmission range of the line-filter-group high-pass section (in many cases the condition $OL = 20 \log_{10} g \approx 0$ decibels will apply), it is possible to calculate the constant minimum reference return loss RL_{R1} which will satisfy the conditions of this example.

The method of solution adopted in this example illustrates the fact that specification of the reference performance characteristics of equipment units may now become an exact science rather than a process based largely on experience.

(11) RELATIONSHIP BETWEEN REFERENCE PARAMETERS AND IMAGE REFERENCE PARAMETERS

Quadripole performance characteristics are frequently given in terms of the image parameters, and it therefore becomes desirable to establish a relationship between the image reference parameters defined in Section 5.1 and a set of quadripole parameters corresponding to operation under non-image operating reference conditions.

The operating factor $G'_{EE,0}$ of an interacting quadripole operating between impedances Z_S and Z_L has been derived in eqn. (18). Under image reference operating conditions

$$G = G_i, \quad Z_{R1} = Z_{N1} \text{ and } Z_{R2} = Z_{N2},$$

$$\text{i.e. } |A_{R1}| = |A_{R2}| = \infty$$

and eqn. (18) thus simplifies to

$$G'_{EE,0} = G_i \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \sqrt{\left(\frac{A_S^2}{A_S^2 - 1} \frac{A_L^2}{A_L^2 - 1} \right)}$$

If now Z_S and Z_L are considered as nominal source and load impedances, respectively, $G'_{EE,0}$ may be regarded as the reference operating factor of the quadripole under these non-image reference conditions.

The associated input inverse reflection coefficient is readily derived from the formulae obtained in Sections 8 and 10. Thus, at the input of the quadripole,

$$A_j = \frac{A_i A_r - 1}{A_i - A_r} = \frac{A_S A_L G_i^2 - 1}{A_S - A_L G_i^2} = -A_S \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \frac{A_L G_i^2}{A_L G_i^2 - 1}$$

where A_j represents the reference input inverse reflection coefficient. Similarly, the reference output inverse reflection coefficient is determined.

For convenience these results are presented in Table 5, together with formulae applicable to certain special cases.

operating under symmetrical conditions. Use of the appropriate formulae from Table 5 and the curves of Fig. 6 leads to

$$\begin{aligned} \text{Reference operating loss} & \quad \dots \quad 3.0 - 0.35 - 0.55 = 2.1 \text{ dB} \\ \text{Reference input return loss} & \quad \dots \quad 6.6 - 0.55 + 2.7 \simeq 8.7 \text{ dB} \\ \text{Reference output return loss} & \quad \dots \quad 6.6 - 0.55 + 2.7 \simeq 8.7 \text{ dB} \end{aligned}$$

If now the cable is added to a transmission system at a junction having a nominal impedance of 600 ohms, zero angle, and the return losses of the dissimilar junction impedances are 20 dB and 26 dB, respectively, at 3400 c/s against this nominal impedance, use of the appropriate formulae in Table 2 and the curves of Fig. 6 gives

$$VL_{\max} = 0.04 + 0.48 + 0.02 = +0.54 \text{ dB}$$

$$VL_{\min} = -0.04 - 0.51 - 0.02 = -0.57 \text{ dB}$$

$$\text{so that } OL_{EE} + 2.1 - 0.57 \leq OL'_{EE} \leq OL_{EE} + 2.1 + 0.54 \text{ dB}$$

$$\text{or } OL_{EE} + 1.53 \leq OL'_{EE} \leq OL_{EE} + 2.64 \text{ dB}$$

(12) CONCLUSION

It has been shown that introduction of the reference parameters considerably simplifies the solution of problems involving non-reference operation, as may occur in the tandem connection of quadripoles. For calculation purposes, the impedance formulae derived may, with advantage, be converted to formulae expressed in terms of inverse reflection coefficients. The latter formulae may be arranged to have factors of the type $X/(X-1)$, and their loss forms may therefore be readily calculated, since

Table 5

RELATIONSHIP BETWEEN REFERENCE PARAMETERS AND IMAGE REFERENCE PARAMETERS*

Image reference conditions: $Z_{R1} = Z_{N1}$, $Z_{R2} = Z_{N2}$, $G = G_i$.

Reference conditions: Z_S and Z_L considered as nominal impedances.

	Interacting quadripole		Non-interacting quadripole	
	Asymmetrical operation	Symmetrical operation $A_S = A_L = A$	Asymmetrical operation	Symmetrical operation $A_S = A_L = A$
Reference operating factor	$G_i \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \sqrt{\left(\frac{A_S^2}{A_S^2 - 1} \frac{A_L^2}{A_L^2 - 1} \right)}$	$G_i \frac{A^2 G_i^2 - 1}{A^2 G_i^2} \frac{A^2}{A^2 - 1}$	$G_i \sqrt{\left(\frac{A_S^2}{A_S^2 - 1} \frac{A_L^2}{A_L^2 - 1} \right)}$	$G_i \frac{A^2}{A^2 - 1}$
Reference input inverse reflection coefficient	$-A_S \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \frac{A_L G_i^2}{A_L G_i^2 - 1}$	$-A \frac{A^2 G_i^2 - 1}{A^2 G_i^2} \frac{G_i^2}{G_i^2 - 1}$	$-A_S$	$-A$
Reference output inverse reflection coefficient	$-A_L \frac{A_S A_L G_i^2 - 1}{A_S A_L G_i^2} \frac{A_S G_i^2}{A_S G_i^2 - 1}$	$-A \frac{A^2 G_i^2 - 1}{A^2 G_i^2} \frac{G_i^2}{G_i^2 - 1}$	$-A_L$	$-A$

* For convenience the term 'image' is extended to cover non-interacting quadripoles operating under conditions of matched input and output circuits, i.e. $|A_{R1}| = |A_{R2}| = \infty$.

(11.1) Short Length of Voice-Frequency Cable

In Section 5.1.1 the image reference parameters at 3400 c/s of 1 mile of 10 lb/mile voice-frequency cable were quoted. It is desired to determine the reference-loss parameters of this one mile of cable, corresponding to operation between nominal source and load impedances of 600 ohms, zero angle.

The cable may be represented by an interacting quadripole

$20 \log_{10} |X/(X-1)|$ and the angle of $X/(X-1)$ can be obtained from large-scale curves similar to those shown in Fig. 6 or from Tables. If $20 \log_{10} |X|$ and the angle of X are known, precise calculations of return and operating loss are possible; if $20 \log_{10} |X|$ only is known, limits of the loss expressions can be obtained.

The numerical solution of transmission problems by means

of the techniques advanced results in a considerable reduction in labour as compared with other known methods of solution. Further, a clearer understanding of the electrical principles involved is obtained, since, for known variations in the reference characteristics of a quadripole, the effect on its non-reference operation is rapidly seen by inspection of Fig. 6.

Non-reference operation of a quadripole may alternatively be regarded as operation under new reference conditions. Accordingly, by the use of formulae derived it is possible to obtain a new set of reference parameters for the new reference conditions in terms of the reference parameters corresponding to the original reference conditions. In this manner, the image reference parameters have been related to a new set of reference parameters corresponding to selected nominal impedances. Similarly, the iterative reference parameters could be related to a new set of reference parameters.

The choice of quadripole reference-loss parameters simplifies the measurement of quadripole performance characteristics under reference conditions (a fact which is of assistance in production quality control) and permits the minimum reference performance requirements of quadripoles to be specified in such a manner that the loss characteristics of the transmission system formed from their tandem connection meet prescribed limits.

(13) ACKNOWLEDGMENT

The authors wish to extend their thanks to Mr. N. M. Macdonald, of the Engineer-in-Chief's Office, Postmaster-General's Department, Australia, for his interest in the paper during the course of its preparation, and for his suggestions in connection with the determination of loss-expression limits.

(14) REFERENCES

- (1) DØSSING, S., and SEYMOUR, P. W.: 'Loss Characteristics of Tandem Connected Transmission Equipment', *Telecommunication Journal of Australia*, 1953, 9, p. 152.
- (2) LLEWELLYN, F. B.: 'Some Fundamental Properties of Transmission Systems', *Proceedings of the Institute of Radio Engineers*, 1952, 40, p. 271.
- (3) TUCKER, D. G.: 'Two Notes on the Performance of Rectifier Modulators', *Proceedings I.E.E.*, Paper No. 1384 R, November, 1952 (99, Part III, p. 400).
- (4) TUCKER, D. G.: 'Some Aspects of the Design of Balanced Rectifier Modulators for Precision Applications', *Journal I.E.E.*, 1948, 95, Part III, p. 161.
- (5) TUCKER, D. G.: 'Rectifier Modulators with Frequency-Selective Terminations', *Proceedings I.E.E.*, Paper No. 863 R, September, 1949 (96, Part III, p. 422).
- (6) SEYMOUR, P. W.: 'A Method of Analysing the Performance of Tandem-connected Four-terminal Networks', Paper presented at the May, 1955, I.R.E. Radio Engineering Convention, Melbourne, Australia, *Proceedings of the Institution of Radio Engineers (Australia)*, July, 1956.

(15) APPENDICES

(15.1) Interacting Quadripole under Reference Conditions

Let the reference parameters of the quadripole in Fig. 1 be Z_{R1} , Z_{R2} , and G , and let the nominal impedances be Z_{N1} and Z_{N2} . It is then seen that

$$G = \frac{I_0}{I_1} = \left[\frac{V}{2\sqrt{\left(\frac{Z_{N2}}{Z_{N1}}\right)\frac{1}{Z_{N2}}}} \right] \frac{1}{I_1}, \text{ or } I_1 = \frac{V}{2G\sqrt{Z_{N1}Z_{N2}}}$$

Using matrix notation and considering transmission from left to right

$$\begin{bmatrix} V \frac{Z_{R1}}{Z_{N1} + Z_{R1}} \\ V \frac{1}{Z_{N1} + Z_{R1}} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} Z_{N2} & I_1 \\ I_1 \end{bmatrix} \quad (1)$$

Considering transmission from right to left, and applying the reciprocity theorem, which permits the new load current I_2 , and the current I_1 to be related, we obtain

$$\begin{bmatrix} Z_{N1} & I_1 \\ -I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V \frac{Z_{R2}}{Z_{N2} + Z_{R2}} \\ -V \frac{1}{Z_{N2} + Z_{R2}} \end{bmatrix} \quad (2)$$

Using the expression derived for I_1 , then I_1 and V may be eliminated from these matrix equations and the four general circuit constants A , B , C and D can be derived by simple algebra:

$$A = \frac{2\sqrt{(Z_{N1}Z_{N2})G}}{Z_{R2} + Z_{N2}} \left(\frac{Z_{R2} + Z_{N2}}{4Z_{N1}Z_{N2}G^2} Z_{N1} + \frac{Z_{R1}}{Z_{R1} + Z_{N1}} \right) \quad (3)$$

$$B = \frac{2\sqrt{(Z_{N1}Z_{N2})G}}{Z_{R1} + Z_{N1}} Z_{R1} - AZ_{N2} \quad (4)$$

$$C = \frac{2\sqrt{(Z_{N1}Z_{N2})G}}{Z_{R2} + Z_{N2}} \left(\frac{1}{Z_{R1} + Z_{N1}} - \frac{Z_{R2} + Z_{N2}}{4Z_{N1}Z_{N2}G^2} \right) \quad (5)$$

$$D = \frac{2\sqrt{(Z_{N1}Z_{N2})G}}{Z_{R1} + Z_{N1}} - CZ_{N2} \quad (6)$$

(15.2) Ideal Transducer under Non-reference Conditions

From Fig. 16(a):

$$\begin{bmatrix} V - I_3 Z_S & Z_S \\ I_3 \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{Z_{N1}}{Z_{N2}}} & 0 \\ 0 & \sqrt{\frac{Z_{N2}}{Z_{N1}}} \end{bmatrix} \begin{bmatrix} I_4 & Z_L \\ I_4 \end{bmatrix}$$

from which

$$I_3 = \sqrt{\left(\frac{Z_{N2}}{Z_{N1}}\right)} I_4 \quad (7)$$

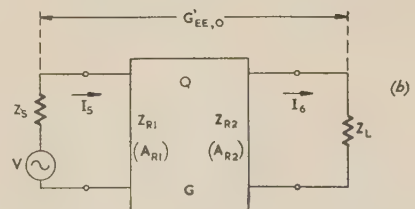
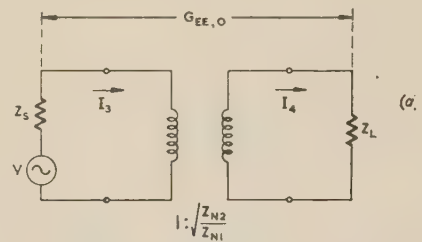


Fig. 16.—Replacement of an ideal transducer with an interacting quadripole.

- (a) Ideal transducer operating between Z_S and Z_L .
(b) Interacting quadripole operating between Z_S and Z_L .

and

$$V - I_3 Z_S = \sqrt{\left(\frac{Z_{N1}}{Z_{N2}}\right) Z_L} I_4 \quad . \quad . \quad . \quad (8) \quad \text{Since}$$

Combining eqns. (7) and (8),

$$I_4 = \frac{V}{\sqrt{(Z_{N1} Z_{N2}) \left(\frac{Z_S}{Z_{N1}} + \frac{Z_L}{Z_{N2}}\right)}} \quad . \quad . \quad . \quad (9)$$

(15.3) Interacting Quadripole under Non-reference Conditions

From Fig. 16(b):

$$\begin{bmatrix} V - I_5 Z_S \\ I_5 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_6 Z_L \\ I_6 \end{bmatrix}$$

from which $I_5 = (C Z_L + D) I_6 \quad . \quad . \quad . \quad (10)$

and $V - I_5 Z_S = (A Z_L + B) I_6 \quad . \quad . \quad . \quad (11)$

Combining eqns. (10) and (11),

$$I_6 = \frac{V}{(A Z_L + B) + (C Z_L + D) Z_S} \quad . \quad . \quad . \quad (12)$$

The operating factor of the ideal transducer under non-reference conditions is $G_{EE,0} = I_0 / I_4$, where

$$I_0 = \frac{V \sqrt{Z_L}}{2 Z_L}$$

is the current that would flow in Z_L if fed from the source via an ideal transducer of impedance ratio Z_L / Z_S . Similarly, the operating factor of the interacting quadripole under non-reference conditions $G'_{EE,0} = I_0 / I_6$. Thus

$$\frac{G'_{EE,0}}{G_{EE,0}} = \frac{I_4}{I_6}$$

Further, from a consideration of system linearity,

$$\frac{I_4}{I_6} = \frac{I'_4}{I'_6} = \frac{G'_{EE}}{G_{EE}} \quad (\text{see Fig. 4}),$$

so that

$$\frac{G'_{EE,0}}{G_{EE,0}} = \frac{G'_{EE}}{G_{EE}} = \frac{I_4}{I_6} = \frac{(A Z_L + B) + (C Z_L + D) Z_S}{\sqrt{(Z_{N1} Z_{N2}) \left(\frac{Z_S}{Z_{N1}} + \frac{Z_L}{Z_{N2}}\right)}} \quad (13)$$

Therefore, using eqns. (3)–(6),

$$\frac{G'_{EE,0}}{G_{EE,0}} = \frac{G'_{EE}}{G_{EE}} = G \frac{\left(\frac{Z_S}{Z_{N1}} - 1\right) \left(\frac{Z_L}{Z_{N2}} - 1\right)}{2 \left(\frac{Z_S}{Z_{N1}} + \frac{Z_L}{Z_{N2}}\right)} \left\{ \frac{2 \left(\frac{Z_{R1}}{Z_{N1}} + \frac{Z_S}{Z_{N1}}\right) 2 \left(\frac{Z_{R2}}{Z_{N2}} + \frac{Z_L}{Z_{N2}}\right)}{\left(\frac{Z_{R1}}{Z_{N1}} + 1\right) \left(\frac{Z_S}{Z_{N1}} - 1\right) \left(\frac{Z_{R2}}{Z_{N2}} + 1\right) \left(\frac{Z_L}{Z_{N2}} - 1\right)} - \frac{1}{G^2} \right\} \quad (14)$$

$$G_{EE,0} = \frac{I_0}{I_4} = \frac{1}{2} \sqrt{\left(\frac{Z_{N1} Z_{N2}}{Z_S Z_L}\right) \left(\frac{Z_S}{Z_{N1}} + \frac{Z_L}{Z_{N2}}\right)} \quad . \quad . \quad (15)$$

$$G'_{EE,0} = \frac{G}{4} \sqrt{\left(\frac{Z_{N1} Z_{N2}}{Z_S Z_L}\right) \left(\frac{Z_S}{Z_{N1}} - 1\right) \left(\frac{Z_L}{Z_{N2}} - 1\right)} \left[\frac{2 \left(\frac{Z_{R1}}{Z_{N1}} + \frac{Z_S}{Z_{N1}}\right) 2 \left(\frac{Z_{R2}}{Z_{N2}} + \frac{Z_L}{Z_{N2}}\right)}{\left(\frac{Z_{R1}}{Z_{N1}} + 1\right) \left(\frac{Z_S}{Z_{N1}} - 1\right) \left(\frac{Z_{R2}}{Z_{N2}} + 1\right) \left(\frac{Z_L}{Z_{N2}} - 1\right)} - \frac{1}{G^2} \right] \quad . \quad . \quad . \quad (16)$$

Using appropriate inverse reflection coefficients, eqns. (15) and (16) reduce to

$$G_{EE,0} = \frac{A_S A_L - 1}{A_S A_L} \sqrt{\left(\frac{A_S^2}{A_S^2 - 1} \frac{A_L^2}{A_L^2 - 1}\right)} \quad . \quad . \quad (17)$$

$$G'_{EE,0} = G \sqrt{\left(\frac{A_S^2}{A_S^2 - 1} \frac{A_L^2}{A_L^2 - 1}\right)} \left(\frac{A_S A_{R1} - 1}{A_S A_{R1}} \frac{A_{R2} A_L - 1}{A_{R2} A_L} - \frac{1}{A_S A_L G^2} \right) \quad . \quad (18)$$

(15.4) Impedance and Reflection Coefficient of an Interacting Quadripole under Non-reference Conditions

The input impedance of an interacting quadripole under non-reference conditions is, from Fig. 16(b), $Z_{in} = \frac{V - I_5 Z_S}{I_5}$. Therefore, using eqns. (10) and (11),

$$Z_{in} = \frac{A Z_L + B}{C Z_L + D} \quad . \quad . \quad . \quad (19)$$

By the use of eqns. (3)–(6), this becomes

$$Z_{in} = Z_{R1} + \frac{(Z_{R1} + Z_{N1})^2 (Z_{R2} + Z_{N2}) (Z_L - Z_{N2})}{4 Z_{N1} Z_{N2} G^2 (Z_{R2} + Z_L) - (Z_{R1} + Z_{N1}) (Z_{R2} + Z_{N2}) (Z_L - Z_{N2})} \quad . \quad . \quad . \quad (20)$$

Using eqn. (20),

$$\frac{Z_{in} - 1}{Z_{N1}} = \frac{Z_{R1} - 1}{Z_{N1}} + \frac{\left(\frac{Z_{R2}}{Z_{N2}} + 1\right) \left(\frac{Z_L}{Z_{N2}} - 1\right)}{2 G^2 \left(\frac{Z_{R2}}{Z_{N2}} + \frac{Z_L}{Z_{N2}}\right)} \quad . \quad (21)$$

which, in terms of inverse reflection coefficients, becomes

$$\frac{1}{A_{in}} = \frac{1}{A_{R1}} + \frac{A_{R2}}{(A_{R2} A_L - 1) G^2} \quad . \quad . \quad (22)$$

A NEW FORM OF ELECTROLYTIC TANK

By K. F. SANDER, Ph.D., and J. G. YATES, M.A., Associate Members.

(The paper was first received 24th March, and in revised form 13th June, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

The chief cause of erratic results when field gradients are measured in an electrolytic tank by means of closely spaced probes is shown to be variability of the meniscus near the probes. A new form of tank is described in which measurements are made by capillary probes in a plane defined by an insulating surface, thereby eliminating the meniscus effects present when the probes are placed in a free surface.

(1) INTRODUCTION

In the normal type of electrolytic tank, measurements are made by inserting one or more probes in the free surface. The potential differences between the probes may be used to obtain an approximation to the field gradient at the point. Such measurements are liable to errors of two kinds.

First, the finite spacing between the probes and their finite size mean that a p.d. between two probes is not equal to the field strength in that direction multiplied by the spacing. In fields which are not changing too rapidly compared with probe spacing, the relation between p.d. and field strength is one of proportionality, with the constant representing effective spacing. This effective spacing is a function of depth of penetration of the probes.

Secondly, erratic errors exist which are due to fluctuations in the level of the free surface and to inhomogeneity of the electrolyte. The fluctuations in the free surface affect the p.d. between probes in two ways. The less serious of these is that due to the variation of mean level due to evaporation, which will alter the effective spacing between the probes. This level can be maintained to any desired constancy by use of a suitable depth gauge. The more serious is the variation of meniscus associated with an array of closely spaced probes. Such a meniscus will depend greatly on the state of cleanliness of the surfaces of both liquid and probe, and will in general be highly erratic.

The results of measurements on the effect of meniscus variation are discussed in the paper. A modification to normal-type probe arrays is suggested to reduce erratic errors due to meniscus variation. Further, a novel type of tank is described in which the free surface is replaced by a flat surface completely submerged and therefore meniscus-free.

(2) MEASUREMENTS ON EFFECT OF MENISCUS

With a single isolated probe, the effect of surface tension is to raise the surface level in the immediate neighbourhood of the probe. This, in practice, produces a negligible error of the potential measured by the probe. If a second probe is placed near the first, each probe will continue to measure the potential of the point which it occupies, but the potential field will be perturbed by the disturbance of the electrolyte near the probes. This perturbation is much greater than that at a corresponding point near an isolated probe because of the general raising of surface level between the probes, as indicated in Fig. 1. This

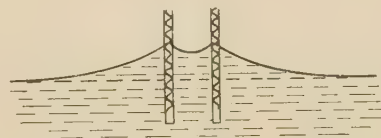


Fig. 1.—Meniscus in the neighbourhood of a two-probe array.

extra bulk of liquid in the immediate neighbourhood of the probes can be regarded as effectively increasing the electrolyte conductivity in that neighbourhood and hence giving rise to a measured field gradient below the true value. It had been found in the use of an electrolytic tank with arrays of four probes that erratic variations of several parts per cent in the measured field gradient could take place under unfavourable conditions. If these are due to variations in the meniscus, the reduction of the field gradient caused by the meniscus must be several times greater, i.e. about 10% of the p.d. between probes.

Two investigations have been carried out: first on a scaled-up model with artificial menisci, and secondly on an actual system of probes in which arrangements were made to observe the meniscus behaviour simultaneously with the p.d. between probes. The results of these are reported below.

(2.1) Measuring System

The measuring system used was essentially that described by the authors in a previous paper,¹ except for one change. The measurements were made with a specially designed difference amplifier in which the p.d. between a pair of probes, each of which was connected to the grid of a valve, is compared with a voltage produced from a potentiometer applied between the cathodes of the valves, adjustment being made for a null output. The resistive potentiometer used earlier has now been replaced by a tapped transformer of low leakage. The design of such a transformer has been considered elsewhere,² and it has been shown that it is easy to design a suitable transformer on a small Mumetal core having errors due to leakage of less than 1 in 10⁴. The present system uses a transformer with three windings. One winding tapped every ten turns and another every turn provide settings to 1%. A slide-wire tapped across the third winding (of two turns) gives intermediate positions. This is a simple, cheap and reliable arrangement which gives readings accurate to within 1 part in 10⁴ of full scale. All the measurements are by this means made directly as a fraction of the total voltage applied to the tank electrodes.

(2.2) Measurements on a Scale Model

Arrays of probes which the authors have used have been made of glass capillaries about 0.5 mm in diameter, separated by 3 or 4 mm. It was decided, for reasons of mechanical convenience, to make the scale probes about $\frac{1}{4}$ in in diameter. This corresponds to a scale of 3 mm to 1 in. The scale probes are therefore separated by 1 in. The model used is shown in Fig. 2. It consisted of a cubical box of about 5 in side, open at the bottom, with two opposite sides of stainless-steel sheet and

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Sander and Mr. Yates are in the Engineering Department, University of Cambridge.

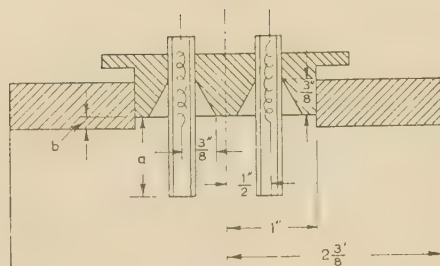


Fig. 2.—Section through model tank and probes for investigating effect of meniscus.

the remaining three sides of Perspex. In the top a hole was cut to enable different probe sets to be mounted easily. The block mounting the probes was normally flush with the under surface of the top, although it later proved useful to raise it, as indicated in the diagram. The menisci were simulated by conical holes $\frac{3}{8}$ in in diameter at base and of 90° apex angle. The effect of these menisci could be eliminated by slipping mica discs over the probes, of diameter sufficient to cover the bottom of the conical hole. This left a substantially plane surface on the underside of the top. The whole box stood in a dish so that its interior was full of electrolyte. The under-surface of the top then simulated probes in the free surface of an electrolyte. A p.d. was applied between the steel sides to create a uniform field across the box.

Measurements were made on both capillary probes and probes of solid metal.

(2.2.1) Measurements with Capillary Probes

The probes were of Perspex, $\frac{1}{4}$ in in diameter with a $\frac{1}{8}$ in hole in the middle. Contact was made with an electrode about $\frac{1}{2}$ in from the opening. The variables were the dimensions a and b shown in Fig. 2. The measurement of p.d. between probes made with probe tips flush with surface and with mica discs covering the conical holes agreed with that calculated from the dimensions of the structure to an accuracy better than that of the mechanical measurements. An effective spacing for any other set of probes is defined as that distance which when multiplied by the true field gradient will give the measured p.d. The ratio of effective to actual spacing is obtained directly as the ratio of measured p.d. to the p.d. in this particular case measured when the probes are flush with the plane surface.

In Table 1 are given values of effective spacings of capillary probes of varying penetration, with and without the artificial menisci.

Table 1

Penetration (a) when $b = 0$ (Fig. 2)	Effective spacing	
	With plane surface	With artificial menisci
in	in	in
0	1.000	0.993
0.5	1.021	1.017

The errors as shown by Table 1 are too small to account for the effects noticed. An attempt was therefore made to simulate the effect of the raising of the surface level near the probes by altering dimension b in Fig. 2. The results obtained are given in Table 2.

Table 2

Penetration (a) (Fig. 2)	Effective spacing with plane surface	
	when $b = 0$ (Fig. 2)	when $b = 1/16$ in
in	in	in
0	1.000	0.967
0.25	1.012	0.986
0.50	1.021	1.002

The difference between the two columns is of the order of 2%. The raising of level by $\frac{1}{16}$ in, which is one-quarter of the probe diameter, corresponds to about one-quarter of a millimetre in an actual array. Casual observations of the level of the meniscus near a probe set in a tank indicate a possible raising of this level above the free surface by 1 mm. The experiment therefore indicates that the variations in meniscus may account for the erratic fluctuations observed in practice.

(2.2.2) Metal Probes.

The capillary probes were replaced by stainless-steel probes of the same diameter. Results, which are similar to those in Table 2, are given in Table 3. The conical hole gave effects of the same magnitude as with capillary probes.

Table 3

Penetration (a) (Fig. 2)	Effective spacing with plane surface	
	when $b = 0$ (Fig. 2)	when $b = 1/16$ in
in	in	in
0	0.991	—
0.25	0.981	0.950
0.50	0.969	0.943

It may be remarked that the depth sensitivities of the capillary and metal probes are opposite in sign, as might be expected, and nearly equal. The actual perturbation due to the insertion of the probes is not greatly different for the two kinds. At zero penetration a 1% change in effective spacing is observed for the metal probes.

It therefore appears that, as claimed in Reference 1, the perturbation of the capillary probe in the surface is much less than that of a metal probe, but since in present practice penetration must take place, the mass of liquid displaced produces a much larger perturbation. Hence this advantage of capillary probes cannot be realized with present systems.

(2.3) Measurements on an Actual Probe Array

For measurements on a probe array it was necessary to be able to examine the shape of the liquid surface near the probe at the same time as electrical measurements were being made. A small Perspex tank was made, with two opposite sides of stainless steel which were 2 in apart. This tank could be filled to the brim so that the free surface of the liquid was slightly above the tank sides. A graduated microscope was used to examine the meniscus produced round a pair of probes. By focusing on the surface away from the probes, an estimate of the free surface position could be made. An array of two probes, consisting of glass capillaries about 0.5 mm in diameter separated by a distance of 4 mm was used, mounted kinematically so that it could be replaced in exactly the same position. The

probes were placed in the direction of the electric field and at right angles to the line of sight of the microscope. Measurements were made of the height of the surface between the probes above the free surface level and of the p.d. between the probes. The meniscus was changed by removing and replacing the probes, or by disturbing the liquid in their neighbourhood by means of a piece of wire. It was found possible to produce variations of over 1 mm. The results showed a high degree of correlation between height of meniscus and p.d., as evidenced by the curve in Fig. 3. The range of apparent field values shows

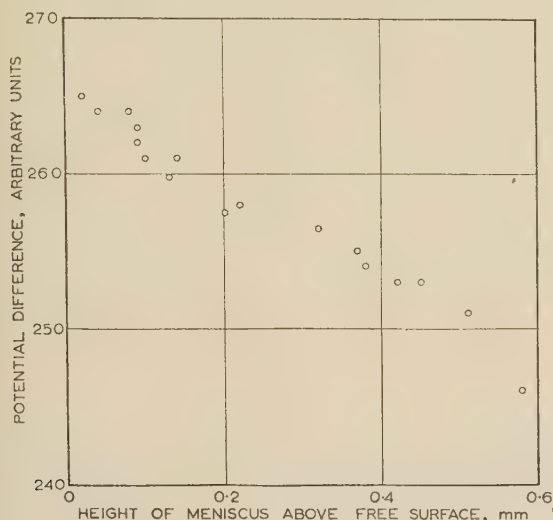


Fig. 3.—Variation of p.d. between probes with height of meniscus.

a variation of nearly 10% between the extremes. It therefore appears that the conclusions of the somewhat artificial experiment described in Section 2.2 were justified, and that the variability of meniscus height can cause relatively great changes in p.d. between the probes.

(2.4) Conclusions on the Design of Probe Arrays

It appears from the foregoing measurements that the variation of meniscus form is one of the chief causes of error in precision measurements of gradient in electrolytic tanks. One method of overcoming this limitation is to replace the free surface by a closed one in which the shape of the surface is controlled. This has been suggested by the authors¹ and a practical tank based on this idea is described in Section 3. However, apart from this, some improvement can be made by suitable design of probe arrays. The purpose of any improvement must be to isolate the meniscus from the point where measurements are made. This can be done, as reported by Hollway,³ by insulating metal probes and forming them into one support, so that the level between the probes is fixed by the shape of the support. This method has various disadvantages: the system becomes bulky, and the depth of penetration is liable to be excessive for measurements near the axis of an axially symmetrical system. It also causes a high sensitivity to depth variations. An alternative way is to use a system such as that illustrated in Fig. 4. In this diagram M is a mica disc attached to the tips of the capillary

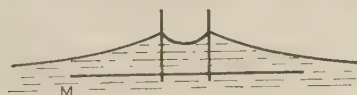


Fig. 4.—Modified two-probe array.

probes. In a two-dimensional field a very thin insulating layer of this kind parallel to the surface would not affect the field distribution. The perturbations due to the meniscus will be attenuated before reaching the probe tips, the attenuation becoming greater as the disc is enlarged. Such an array suffers from disadvantages if it is required to approach an electrode; also, the depth of the disc must be small when axially symmetrical fields are being measured. However, the rise of surface level is compensated to some extent if the depth is chosen correctly. On account of the meniscus near an electrode, measurements of gradient in such a position would in any case be inaccurate. Such a probe set was made from a capillary glass array of dimensions similar to those already given by adding a $\frac{1}{2}$ in-diameter mica disc with holes to allow the glass probes to penetrate it. The diameter of this disc is about three times the probe spacing. Scale measurements on this array, similar to those given in Section 2.2, showed a variation of 1% over the same range of meniscus heights. It would therefore appear to have some merits compared with the normal type.

The authors have previously suggested a figure of 0.1% as a working possibility in a normal-type tank, whereas the figures given in Sections 2.3 and 2.4 appear to indicate considerably increased error. In practice, the tank has been used as an element in an automatic electron trajectory plotter, in which the probes are in movement and the free surface is continually disturbed. An initial settling-down process lasting several hours has been observed, after which the variations are much closer to the 0.1% level.

(3) THE INVERTED TANK

In the authors' earlier paper,¹ it was suggested that the advantages of capillary probes would be exploited to the full if they could be used in the form of small holes in an insulating sheet placed in the surface of symmetry of the system. Measurements reported in that paper indicated the great stability to be expected from such a system.

Two ways of realizing a system of this type can be imagined. One is that referred to above of an insulating sheet sliding on top of the electrode system of a normal tank, so that the sheet replaces the normal free surface. The other is to invert the positions of surface and model, making the bottom of the tank flat with a fixed set of hole probes, and moving the model relative to the probes. The choice between the two possibilities is affected by the fact that the first requires a total tank of linear dimensions three times those of the model, whereas the second only requires a factor of two. The tank will necessarily be larger than the usual type with probes in the free surface. In either case the flat surface has to be maintained rigid to a predetermined order of accuracy, and hence must be fairly heavy. There seems little difference in difficulty between moving the rigid plate and moving the model, and it was therefore decided to make a tank based on the second method.

(3.1) Description of Tank

A tank has been constructed to investigate the possibilities of the method, and has proved very successful. It has overall dimensions of 3 ft 6 in \times 2 ft \times 4 in. The small depth precludes its use for half-models of axially symmetrical systems, but this is a matter of detail only. An angle-iron frame was constructed, rigid enough to enable distortions to be kept to within a few thousandths of an inch. The bottom of the tank was made of $\frac{1}{4}$ in plate glass, in the centre of which a hole 1 in diameter was cut to mount a Perspex probe assembly. It was arbitrarily decided to aim for the bottom to be planar to within 0.005 in. In order to hold the glass flat to this tolerance, it was rested on

sixteen 2 B.A. screws tapped into the angle-iron frame. These were placed in four rows spaced equally from edge to edge. A cross-section of the bottom through the probe array is shown in Fig. 5. An insert was made of Tufnol with a reamed hole

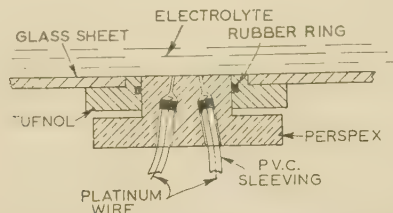


Fig. 5.—Section through probe system of the inverted tank.

to take probe sets of standard diameter. A rubber O-ring was included to make the assembly watertight. The insert was glued into position with a cold-setting cement, the top surface being slightly proud of the glass, and the surface was then rubbed down to be co-planar with the glass.

The probe assembly was built into a Perspex rod fitting the hole in the Tufnol insert. Four probe holes were drilled, initially about 1 mm in diameter. They were inclined to the normal in order to increase the distance between the exit holes, and thus increase the resistance of leakage paths. Contact was made by platinum wire forced through Neoprene bungs tightly held in the wider portion of the hole. The length of leakage path was further increased by p.v.c. sleeving covering the wire.

The sides of the tank were also of plate glass, cemented to the bottom. This has not proved entirely satisfactory, and it is proposed to use a flexible joint between sides and bottom in future tanks.

(3.2) The Carriage

The carriage supporting the model has to maintain height to within 0.005 in. Two ground steel rods running the length of the tank, are supported at each end and in the middle; these have to be adjusted to be straight to within 0.001 in. A carriage runs on the rods, and, since they are straight and parallel to a high degree of accuracy, it is mounted at four points. On one side are two sets of wheels (actually, small ball-races) shown in Fig. 6, and on the other side are two sets, each with a

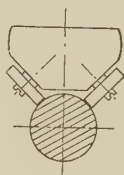


Fig. 6.—Runner of model carriage.

single race, running on the crown of the rod. In this way difficulties associated with the rigidity of the carriage under torsion are avoided.

The carriage itself has two cross-runners, similar in design to the longitudinal runners but much shorter, and these carry a trolley mounted with similar ball-races. Because of its small size the trolley is mounted kinematically with only one race on the crown of the rod. Models are mounted from the trolley, again kinematically, so that replacement of one model by another is easy and reversible without alterations.

(3.3) Alignment Procedure

After some trial and error, a simple and satisfactory process was found for aligning the tank bottom and runners to the

required degree of accuracy. Alignment of the runners was first carried out, as follows. One runner was compared with a straight-edge using feeler gauges, and adjusted, by means of the centre support, to be straight to within 0.001 in. The straight-edge was then used across the diagonals of the tank, and the ends of the other runner were adjusted until the centre of each diagonal was the same height above an arbitrary base. This ensured that the ends of the runners were in a plane. The centre support of the second runner was then adjusted to make the second runner straight to within 0.001 in. This process finished with the two runners straight and parallel. With a dial gauge clamped to the straight-edge, the depth of the bottom could then be measured with respect to the plane of the runners. The height of the glass above each screw in succession was adjusted to a constant value, and a further check was then made for final adjustments. In this way the bottom was easily made flat to within 0.005 in. Loading with water (about 1 cwt) produced a change of 0.001 in in the centre of the tank. The flatness was checked after six months, and again after the glass had been removed and the sides re-cemented, without touching the adjusting screws. In each case the changes observed on checking were of the order of 0.001 in. It was therefore concluded that the mechanical stability of the structure is entirely adequate.

(3.4) Techniques in the use of the Inverted Tank

(3.4.1) Probes.

The form of surface probes described has two important advantages over the usual type. First, they are unbreakable, and no form of limit switch is required to avoid collision between probes and model. Secondly, they can be made much more precisely. Errors in position of the holes in four probe arrays will give different scale factors in measurement of gradient in two directions, and will further imply that the measured components of the gradient are not referred to perpendicular axes. In practice these errors can be compensated electrically, but the requirements for stability of the electrical networks involved are much easier to satisfy when the errors are small. Because of the elimination of meniscus variations, much closer spacings can be used with advantage. The spacing can, in fact, be reduced to the order of the hole diameters. Whilst the hole diameter has an effect on the effective spacing, scale measurements have shown that with the holes almost touching it is only 4% less than the distance between centres. In practice the effective spacing is always determined by electrical calibration.

The reduction of probe spacing may, in turn, allow the use of smaller models.

(3.4.2) Models.

Because the model moves in relation to the tank walls, it must be completely enclosed, so that the field is unperturbed by the tank walls. Closing of the top ensures the complete absence of any effect due to ripples on the free surface.

The models differ slightly from the usual mode of construction in that they must be true half-models, with no projections which would prevent them coming into close contact with the tank bottom. So far as mounting is concerned, the changes required have proved to have definite advantages.

Electrically, there is one significant difference: in the inverted tank it is impossible to make a complete break between two regions of electrolyte. When the boundary is metallic, as it most often is, this is unimportant, but it is entirely otherwise with insulating boundaries, where the leakage through the gap between the boundary and the tank bottom may be serious. This difficulty can be overcome in various ways, depending on the problem. For example, the use of a simple system of guard

rings will give the correct field everywhere except in the immediate vicinity of the boundary. Measurements near the centre of a parallel-plate condenser showed negligible errors even with gaps several times the normal maximum.

(4) OVERALL CHECK OF ACCURACY

In order to obtain an overall check of errors from all sources, measurements were made on the field between concentric cylinders, this system being chosen because the model can be made with high precision. The inner cylinder consisted of 1 in-diameter ground stainless-steel rod, and the outer cylinder of $\frac{1}{4}$ in-thick dull nickel-plated brass, which was bored out before plating to an inside diameter of 15.0 cm. The two elements were mounted on a $\frac{1}{2}$ in-thick Perspex sheet with rings turned in it to locate them accurately. The system was about 2 in deep. The eccentricity amounted to less than 0.1 mm. This model was mounted on a support which was located kinematically on to the tank carriage, provision being made for height adjustment by a micrometer at each point of support. The model was completely immersed in the electrolyte.

The particular probe set used to make measurements on this model had each pair of probes separated by 0.508 cm between the centres and each probe hole was 0.14 cm in diameter.

(4.1) Experimental Procedure

The object of the experiment was to show that results could be obtained repeatedly to within 0.1% of maximum gradient, both over long periods and after withdrawing and replacing the model in the tank. Since this latter check is partially one of the goodness of design of the kinematic mount, a procedure was adopted which would separate effects due to defects in the design of the mount and defects in other parts of the system. This procedure was to measure the field gradient along a diameter of the model, at sets of points equally spaced on either side of the centre. Differences between readings at identical points on opposite sides of the centre are then a measure of the accuracy of setting of the centre position, whilst averaging the readings at such pairs of points effectively eliminates errors of positioning. The actual positions were read from a millimetre scale observed through a microscope with a scale on the eyepiece. In this way positions were reset to ± 0.01 mm.

(4.2) Stability

It was found that circulation of the electrolyte in the tank and through the model was essential to obtain short-term stability of measurement of 0.1% of the gradient. With this precaution the stability from day to day was found to be no worse than over a few minutes. Withdrawal and replacement of the model showed that the kinematic mount, in fact, permitted errors of about 0.1 mm in the position. The effect of these was averaged out, as already mentioned, and the averages showed no significant change due to such withdrawals.

(4.3) Results

The p.d. between the probes was measured for radii between 2 and 7 cm on both sides of the centre. The averages of each pair of these results are given in Table 4. With these are given the theoretical p.d. between probes, calculated from the dimensions of the model, positions of the probes, and details of the electrical measuring system. (This p.d. differs slightly from the product of distance between probes and field gradient at the mid-point between the probes, because of the non-uniformity of the field. At the smallest radius the difference is 0.5%.) Because

the effective spacing between the probes is not equal to the distance between centres, the observed and theoretical values can best be compared by calculating their ratio. The constancy or otherwise of this ratio will be a measure of the accuracy. This is given in the last column of Table 4.

Table 4

Radius	P.D. between probes in arbitrary units		Effective spacing Actual spacing
	Measured	Calculated	
cm			
2.0	113.78	115.00	0.9894 ± 0.0004
2.5	90.94	91.83	0.9903 ± 0.0005
3.0	75.67	76.45	0.9898 ± 0.0005
3.5	64.85	65.49	0.9903 ± 0.0008
4.0	56.75	57.28	0.9909 ± 0.0009
4.5	50.40	50.90	0.9904 ± 0.0010
5.0	45.35	45.80	0.9900 ± 0.0012
5.5	41.22	41.63	0.9901 ± 0.0012
6.0	37.79	38.16	0.9905 ± 0.0013
6.5	34.89	35.22	0.9905 ± 0.0014
7.0	32.22	32.70	0.9853 ± 0.0015

Gap between model and bottom of tank, 0.6 mm.
Tank p.d. 3 V peak, equal to 800 units of p.d.

(4.4) Analysis of Results

Apart from the first and last readings, the effective spacing is very constant. The average is 0.9903 and hence the spacing is constant to within the estimated accuracy of the measurements. This figure agrees with measurements on scale models. The small value of the first and last readings is explained by the finite gap between model and tank bottom. With the circulating system used, the electrolyte was pumped through this gap, and as a result it was difficult to use small spacings continually. However, reduction of the gap to 0.1 mm raised the last reading to 32.35 units. This figure gives a ratio of 0.9893, which compares more favourably with the average of 0.9903. A similar explanation holds for the value at radius 2 cm. Analysis of the two-dimensional system obtained by letting the inner and outer radii become infinite, indicates that the field should be within 0.1% at ten gap-widths away from the electrode face. The measurement in Table 4 was made with one probe not more than four gap-widths away, so that the errors are of the correct order of magnitude to be accounted for by this cause.

A further possible cause of error is the accuracy of mechanical setting. An estimate of the constancy of resetting to ± 0.05 unit of p.d. can be obtained from the first differences of the p.d. between probes. The resetting requires to be done to within 0.01 mm at 2 cm radius and to within ± 0.1 mm at 7 cm radius. The higher tolerance was barely achieved with the system used.

(5) CONCLUSIONS

The mechanical and electrical stability of the inverted tank described has proved extremely satisfactory. A great merit of the probe system described is that once a particular set is installed and calibrated, the calibration will hold for an indefinite period. The mechanical construction is such that the possibility of damage is very small.

(6) ACKNOWLEDGMENTS

The authors wish to acknowledge the help of Mr. R. M. Hill, who made the measurements given in Section 2.3 and designed the disc-type probe described in Section 2.4. They also wish to

acknowledge the assistance given by Mr. A. Barker during the construction of the new tank. Thanks are due to the Admiralty for defraying the cost of apparatus.

(7) REFERENCES

- (1) SANDER, K. F., and YATES, J. G.: 'The Accurate Mapping of Electric Fields in an Electrolytic Tank', *Proceedings I.E.E.*, Paper No. 1370 M, April, 1953 (**100**, Part II, p. 167).
 - (2) OATLEY, C. W., and YATES, J. G.: 'Bridges with Coupled Inductive Ratio Arms as Precision Instruments for the Comparison of Laboratory Standards of Resistance or Capacitance', *ibid.*, Paper No. 1631 M, March, 1954 (**101**, Part III, p. 91).
 - (3) HOLLWAY, D. L.: 'An Electrolytic-Tank Equipment for the Determination of Electron Trajectories, Potential and Gradient', *ibid.*, Paper No. 1837 M, May, 1955 (**103**, Part B, p. 155).
-

THE DIFFRACTION OF AN ELECTROMAGNETIC WAVE BY A CIRCULAR APERTURE

By R. F. MILLAR, M.A.

(The paper was first received 10th March, and in revised form 24th April, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

The effects of interaction across a circular aperture in a plane, perfectly conducting screen are studied by application of the approximate edge-current diffraction theory previously developed,¹ the interaction terms being found by asymptotic evaluation of the first-order aperture field. Their effect is examined by computing the aperture and axial fields and comparing the results with exact theory. An interaction correction to the geometrical-optics transmission coefficient, which agrees well with rigorous theory except for a monotonic component, is obtained. With an assumed current distribution over the entire screen, an expression is found for the transmission coefficient, neglecting interaction, which exhibits monotonic behaviour at variance with that expected from exact theory. A term with the desired variation is derived empirically.

The possibility of extending the method to apertures of more general form is briefly mentioned.

LIST OF SYMBOLS

- $A_1(\eta), A_2(\eta),$
 $A_3(\eta), A_4(\eta)$ = Functions of $J_n(\eta)$ and $N_n(\eta)$ for $n = 0, 1, 2$.
 a = Radius of circular aperture.
 c = Velocity of light in free space.
 ds = Element of rim of aperture at Q .
 $E(x, y, z)$ = Total electric field.
 E_x, E_y, E_z = Components of $E(x, y, z)$.
 E_x^s, E_y^s = Scattered components of E_x, E_y .
 E_x^{int}, E_y^{int} = Interaction components of aperture field in directions of increasing r and θ_p , respectively.
 $E^i(z)$ = Incident electric field of unit intensity.
 E_Y^i = Incident electric field parallel to Y -axis.
 E_Y^s = Electric field scattered by half-plane with E_Y^i incident.
 E_y^0 = Non-interaction component of E_y .
 E_y^i = Interaction component of E_y .
 $F(x)$ = Complex Fresnel integral of argument x .
 $H_x(x, y, z)$ = x -component of total magnetic field.
 $H_0^{(2)}(\eta)$ = Zero-order Hankel function of the second kind and argument η .
 I^{int} = Edge current corresponding to E_y^s .
 I_1, I_2 = Integrals involved in the evaluation of t^0 .
 J = Current density on screen.
 $J_n(\eta)$ = Bessel function of the first kind of order n and argument η .
 J_x, J_y = Components of J .
 k = ω/c .
 $N_n(\eta)$ = Bessel function of the second kind of order n and argument η .
 O = Origin of co-ordinates at centre of circular aperture.
 P = Point in circular aperture.
 Q = Point on rim of aperture.

R = Distance between P and Q ; also distance between edge of half-plane and point of observation.

(r, θ_p) = Polar co-ordinates of point P in aperture.

T = Point on axis of aperture.

t = Transmission coefficient of circular aperture for normal incidence.

t^0, t' = Components of t corresponding to E_y^0, E_y' , respectively.

$(x, y, z), (X, Y, Z)$ = Rectangular Cartesian co-ordinates.

$Z_0 = \sqrt{(\mu_0/\epsilon_0)}$.

α, β = Complex functions of ka .

γ = Angle between positive x -axis and PQ .

δ = Empirically determined constant.

ϵ_0 = Permittivity of free space.

$\eta = ka/2$.

θ = Angle between $E^i(0)$ and ds .

λ = Wavelength of incident radiation.

μ_0 = Permeability of free space.

$\Pi_y(x, y, z)$ = y -component of total Hertz vector.

$\Pi_y^s(x, y, z)$ = y -component of Hertz vector of field scattered by J .

$\Pi_y^i(z)$ = Hertz vector of incident field.

ρ = Distance between P and Q , normalized to $a = 1$.

$\phi = \theta - \theta_p$.

$\chi = \gamma - \theta_p$.

ψ = Angle between aperture plane and plane containing ds and TQ .

ω = Angular frequency of incident radiation.

(1) INTRODUCTION

In a recent Monograph¹ an approximate method was developed for determining the field diffracted by an aperture in a plane, perfectly conducting screen of vanishing thickness. The approximation was achieved by assuming the aperture to be the envelope of a system of perfectly conducting half-planes lying in the plane of the screen. The orientation of each half-plane was thus determined by the direction of the tangent at the corresponding point on the aperture rim. Since the far field scattered by any of these half-planes could be regarded (with certain restrictions) as the radiation field arising from fictitious currents on its edge, the field was given in terms of line integrals taken around the aperture rim.

The edge currents were derived on the assumption that the corresponding half-plane was isolated in space and excited only by the incident field. While this appears to be a valid first approximation, the interaction excitation from currents flowing in other parts of the screen should not be entirely ignored.

The purpose of the present paper is to extend the former theory to the consideration of first-order interaction effects in the particular case of the diffraction of a normally incident plane wave by a circular aperture. The form of the integrals for the aperture field [eqns. (8) of Reference 1] suggests asymptotic evaluation in inverse powers of $(ka)^{1/2}$ (where $k = 2\pi/\lambda$) by the method of steepest descents for points not too near the centre

Correspondence on Monographs is invited for consideration with a view to publication.
 Mr. Millar is at the Cavendish Laboratory, Cambridge, on leave of absence from the National Research Council of Canada.

of the aperture. The field at a point near the rim is then found to consist of two parts. One may be interpreted as arising from currents flowing in the screen near the point, while the other represents the interaction field from currents in the remainder of the screen. The latter is, to the first order, a plane wave at grazing incidence from across the aperture, and its effect is readily calculated when once more the neighbouring part of the screen is taken to be a properly oriented half-plane.

For any given ka , when interaction is taken into account in this way, the scattered fields in the aperture far from the edge, and on the axis, are found to be merely the corresponding fields when interaction is neglected multiplied by a (complex) constant. Near the rim a more elaborate relationship exists.

The transmission coefficient may be determined from knowledge of the axial far-field. It is possible to calculate the component of the total far-field due to interaction from the corresponding edge-current, since in this case it is only near and on the screen that the concept becomes invalid. On the other hand, the non-interaction axial far-field must be found by some other method, for in that case the edge-current concept is not applicable. To calculate this component it is necessary to postulate the currents flowing at all points on the screen when interaction is neglected. The currents chosen, together with justification for their selection, will be considered subsequently.

By employing the variational principles of Levine and Schwinger,² two different expressions for the transmission coefficient of a large circular aperture have recently been obtained by Chang.³ One results from an approximate solution for the current density on the screen, while the second corresponds to a similar solution for the tangential electric field in the aperture. Of the two, the screen-current formulation appears to compare more favourably with the results of rigorous theory. Levine⁴ has obtained corresponding results for the scalar problem of diffraction by a circular aperture at high frequencies. The transmission coefficient yielded by the following analysis is identical to the better of Chang's approximations to order $1/(ka)^{3/2}$.

(2) DEFINITIONS AND NOTATION

The same definitions and notation as employed in Reference 1 will be used when possible, and for convenience they are recapitulated here (see also Fig. 1). The diffracting aperture and plane screen occupy the plane $z = 0$ in an orthogonal system of Cartesian co-ordinates, the origin being taken at the

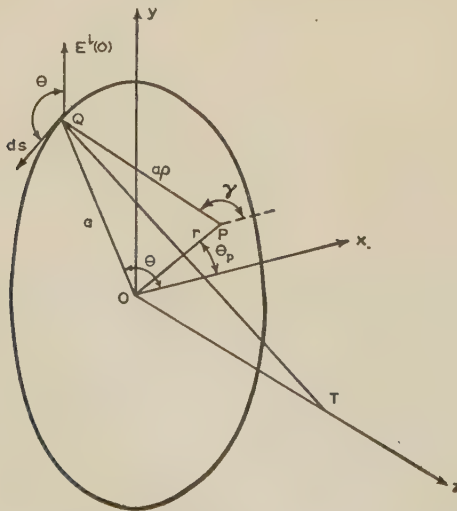


Fig. 1.—Circular aperture in a plane screen, with associated symbols.

centre of the circular aperture. An incident plane wave falls normally upon the screen and aperture from the negative z -direction with z -dependence e^{-jkz} . The propagation coefficient k is equal to $2\pi/\lambda$. Rationalized M.K.S. units are adopted and time dependence $e^{j\omega t}$ is suppressed for all field quantities. Impedance is defined in terms of the permittivity and permeability of free space by $Z_0 = \sqrt{(\mu_0/\epsilon_0)}$. The incident and total electric fields are denoted by $E^i(z) = E^i(0)e^{-jkz}$ [with $E^i(0)$ lying in the direction of the positive y -axis], and $E(x, y, z)$, respectively.

Fig. 1 shows most of the quantities involved in the following discussion. The angle ψ (which is not shown in the Figure) between the aperture plane and the plane containing ds and TQ is greater than or equal to π for points on the 'shadow' side, i.e. $z \geq 0$.

(3) THE APERTURE FIELD

Eqns. (8) of Reference 1 for the tangential electric field in the aperture provide the starting-point in the following analysis. Written in a slightly different notation, the scattered components take the following form:

$$\left. \begin{aligned} E_x^s &= \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \frac{e^{-jkap}}{\rho} \cos(\gamma - \theta) \sin \gamma \cos \theta d\theta \\ E_y^s &= -\frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \frac{e^{-jkap}}{\rho} \cos(\gamma - \theta) \cos \gamma \cos \theta d\theta \end{aligned} \right\} \quad (1)$$

$$\text{where} \quad \rho^2 = \left(\frac{r}{a}\right)^2 + 1 - 2\frac{r}{a} \cos(\theta - \theta_p) \quad (2)$$

and (r, θ_p) are polar co-ordinates of the point of observation with $\theta_p = 0$ corresponding to the positive x -axis.

The asymptotic expansions of E_x^s, E_y^s in inverse powers of $(ka)^{1/2}$ to order $(ka)^{-3/2}$ are derived in Section 9. [See eqns. (39), (40).] When r is not too small, or too close to a , the following expressions are obtained for the total field:

$$\left. \begin{aligned} E_x &\simeq \frac{\sin 2\theta_p}{2\sqrt{(\pi)r^{1/2}}} \frac{a}{(ka)^{1/2}} \left\{ \left[\frac{e^{-jk(a-r)-j\pi/4}}{(a-r)^{1/2}} + \frac{e^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \right] \right. \\ &\quad + \frac{1}{2r} \frac{1}{ka} \left[\frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{3/2}} (4a^2 - 3ar + 2r^2) \right. \\ &\quad \left. \left. + \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{3/2}} (4a^2 + 3ar + 2r^2) \right] \right\} + 0 \left[\frac{1}{(ka)^{5/2}} \right] \end{aligned} \right\} \quad (3)$$

$$\left. \begin{aligned} E_y &\simeq 1 - \frac{\cos^2 \theta_p}{\sqrt{(\pi)r^{1/2}}} \frac{a}{(ka)^{1/2}} \left[\frac{e^{-jk(a-r)-j\pi/4}}{(a-r)^{1/2}} + \frac{e^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \right] \\ &\quad - \frac{\cos 2\theta_p}{2\sqrt{(\pi)r^{3/2}}} \frac{a^2}{(ka)^{3/2}} \left[\left(\frac{a-r}{a} + \frac{a}{a-r} \right) \frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{1/2}} \right. \\ &\quad \left. + \left(\frac{a+r}{a} + \frac{a}{a+r} \right) \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{1/2}} \right] \\ &\quad - \frac{1}{2\sqrt{(\pi)r^{1/2}}} \frac{a}{(ka)^{3/2}} [r + a \cos^2 \theta_p] \\ &\quad \times \left[\frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{3/2}} + \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{3/2}} \right] + 0 \left[\frac{1}{(ka)^{5/2}} \right] \end{aligned} \right\} \quad (4)$$

The physical meaning of these expressions may be seen by reference to Fig. 2, which shows a plan of the circular aperture. The stationary points of ap in eqns. (1) are at A and B, and it

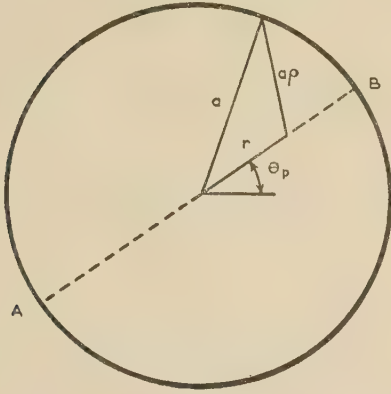


Fig. 2.—Plan of circular aperture, showing the stationary points, A and B, of ap .

follows that the maximum contributions to the integrals are provided by the values of the integrands in the immediate neighbourhoods of these points. The integrals over the remaining portion of the range of integration are small, because of the rapidly oscillating exponential factor in the integrands. The terms in the above expressions involving $a + r$ correspond to A, while those in $a - r$ refer to B.

Evidently the entire scattered field vanishes asymptotically on $\theta_P = \frac{1}{2}\pi, \frac{3}{2}\pi$ to a first approximation, and it is necessary to consider higher-order terms in these regions. Elsewhere, the terms of order $(ka)^{-1/2}$ should provide an adequate approximation if $ka \gg 1$. It should also be noted that these asymptotic expansions are singular at $r = 0$, since in that case ρ has no stationary points.

If the point (r, θ_P) is now allowed to approach the edge of the aperture, certain terms in eqns. (3) and (4) become singular, while others remain finite. The singular terms may be interpreted as the effect of currents flowing near the point and the regular terms represent the interaction from across the aperture.* Only the latter are of interest in the present discussion. Although the form of the singular terms is incorrect, for the edge currents yield only a far-field approximation, the point of observation is in a far-field region as regards the interaction terms. These latter should therefore be given accurately by the finite parts of the above expressions.

(4) THE EFFECT OF INTERACTION

For the remainder of this discussion only the interaction field terms of order $(ka)^{-1/2}$ in eqns. (3) and (4) will be considered. The higher-order terms in the aperture field are not given completely by these equations, which were originally derived from only the first term in the asymptotic expansion of the half-plane field, and their use would be unjustified.

Subject to this restriction, then, the interaction field in the aperture is the following:

$$\left. \begin{aligned} E_x^{int} &\simeq \frac{\sin 2\theta_P}{2\sqrt{(\pi)r^{1/2}} (ka)^{1/2}} \frac{a}{(a+r)^{1/2}} \frac{\varepsilon^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \\ E_y^{int} &\simeq -\frac{\cos^2 \theta_P}{\sqrt{(\pi)r^{1/2}} (ka)^{1/2}} \frac{a}{(a+r)^{1/2}} \frac{\varepsilon^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \end{aligned} \right\} \quad (5)$$

Let $E_X^{int}, E_Y^{int\dagger}$ represent the interaction components in the directions of increasing r and θ_P respectively. Then

* Karp and Russek,⁵ and Clemmow⁶ have considered this type of interaction in the case of an infinite slit between two half-planes.
† The subscripts X, Y are used rather than r, θ_P since this notation proves more convenient at a later stage in the analysis.

$$E_X^{int} \simeq 0,$$

$$E_Y^{int} \simeq -\frac{\cos \theta_P}{\sqrt{(\pi)r^{1/2}} (ka)^{1/2}} \frac{a}{(a+r)^{1/2}} \frac{\varepsilon^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \quad (6)$$

to the order being considered.

If now it is once again assumed that at points in the aperture close to the rim the effect of the adjacent screen is equivalent to that of a properly oriented half-plane, the interaction correction is obtained by the consideration of the two-dimensional problem of the incidence of this field [eqns. (6)] on a half-plane.

The problem of the diffraction of a cylindrical wave by a half-plane was treated many years ago by Carslaw⁷ and Macdonald,⁸ and more recently by Clemmow.⁹ It was shown by Clemmow that, if the line source is far from the diffracting edge, the diffracted wave is, to the first order, the same as that obtaining in the case of diffraction of a plane wave incident from the direction of the cylindrical source. The amplitude and phase of this plane wave at the edge of the half-plane are the same as the corresponding quantities in the cylindrical wave.

The foregoing indicates that the precise variation in amplitude of an incident wave has little effect on the field diffracted by a half-plane, provided only that the source is far enough from the diffracting edge to ensure that the variation of amplitude in the incident wave is not too marked in this region.

The interaction field [eqns. (6)] is, except for the factor $r^{-1/2}$, of this cylindrical nature. Since $r^{-1/2}$ varies slowly near $r = a$, it appears likely that, to the degree of approximation being considered, it is permissible to replace eqns. (6) by the corresponding plane wave.

Consider then a perfectly conducting half-plane occupying $Z = 0, X \geq 0$ in a rectangular Cartesian co-ordinate system (see Fig. 3).

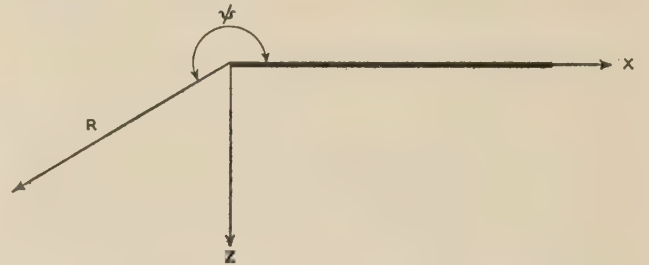


Fig. 3.—Co-ordinate system for half-plane discussion.

The direction of the positive Y-axis is out of the paper.

The incident field is characterized by

$$E_Y^i = -\cos \theta_P \frac{\varepsilon^{-j(2ka-\pi/4)}}{\sqrt{(2\pi ka)}} \varepsilon^{-jkX} \quad (7)$$

The field scattered by the half-plane is, to the first order, at (R, ψ) , the following:

$$E_Y^s \simeq \frac{\cos \theta_P}{2\pi} \frac{\varepsilon^{-j2ka}}{\sqrt{(ka)}} \frac{1}{\sin \frac{1}{2}\psi} \frac{\varepsilon^{-jkR}}{\sqrt{(kR)}} \quad (8)$$

when ψ is not too near 0 or 2π (see, for example, Reference 10, p. 146).

This could be produced by a hypothetical electric current I^{int} flowing along $X = Z = 0$:

$$I^{int} = \frac{\sqrt{2}}{kZ_0} \frac{\varepsilon^{j3\pi/4}}{\sqrt{\pi}} \cos \theta_P \frac{\varepsilon^{-j2ka}}{\sin \frac{1}{2}\psi \sqrt{(ka)}} \quad (9)$$

For the circular aperture, it is now assumed that this same (fictitious) current flows on the rim, from which may be derived the interaction component of the far-field.

(4.1) The Aperture Field

In the aperture, $\psi = \pi$, and for points not too near the rim the interaction components of the scattered field are readily seen to be closely related to the first-order field [eqns. (1)], for the form of the corresponding edge currents is similar. For points in the aperture, $I^{int} = \alpha I^e$, where I^e , defined by eqn. (7) of Reference 1, is the source of the first-order field given by eqns. (1):

$$\alpha = -\frac{\varepsilon^{-j(2ka - \pi/4)}}{2\sqrt{(\pi ka)}} \quad (10)$$

The total scattered field is thus given by the following expressions:

$$\left. \begin{aligned} E_x^s &= \frac{1 + \alpha}{\pi\sqrt{2}} \int_0^{2\pi} \frac{\varepsilon^{-jka\rho}}{\rho} \cos(\gamma - \theta) \sin \gamma \cos \theta d\theta \\ E_y^s &= -\frac{1 + \alpha}{\pi\sqrt{2}} \int_0^{2\pi} \frac{\varepsilon^{-jka\rho}}{\rho} \cos(\gamma - \theta) \cos \gamma \cos \theta d\theta \end{aligned} \right\} \quad (11)$$

The field near the rim is more complex in nature. It may be assumed to consist of four parts: (a) the normally incident field; (b) the field scattered by an isolated half-plane in the presence of (a); (c) the interaction field from across the aperture; (d) the field scattered by an isolated half-plane in the presence of (c). The exact expressions for the fields (b) and (d) in terms of Fresnel integrals (see Reference 9) must now be introduced. They lead to the following approximations to the tangential electric field near the edge of the aperture:

$$\left. \begin{aligned} E_x &= \frac{\sin 2\theta_p}{2\sqrt{\pi}} \left\{ \frac{\varepsilon^{-jk(a-r)-j\pi/4}}{\sqrt{[k(a-r)]}} + \sqrt{\left(\frac{a}{r}\right)} \frac{\varepsilon^{-jk(a+r)+j\pi/4}}{\sqrt{[k(a+r)]}} \right. \\ &\quad \left. - \frac{j}{\sqrt{(2\pi)}} \frac{\varepsilon^{-j2ka}}{\sqrt{(ka)}} \varepsilon^{jk(a-r)} F\{\sqrt{[2k(a-r)]}\} \right\} \\ E_y &= 1 - \frac{2}{\sqrt{\pi}} \left\{ \varepsilon^{j\pi/4} F\{\sqrt{[k(a-r)]}\} \right. \\ &\quad + \frac{\sin^2 \theta_p}{2} \frac{\varepsilon^{-jk(a-r)-j\pi/4}}{\sqrt{[k(a-r)]}} \\ &\quad - \frac{\cos^2 \theta_p}{2} \sqrt{\left(\frac{a}{r}\right)} \frac{\varepsilon^{-jk(a+r)+j\pi/4}}{\sqrt{[k(a+r)]}} \\ &\quad \left. + \frac{j \cos^2 \theta_p}{\sqrt{(2\pi)}} \frac{\varepsilon^{-j2ka}}{\sqrt{(ka)}} \varepsilon^{jk(a-r)} F\{\sqrt{[2k(a-r)]}\} \right\} \end{aligned} \right\} \quad (12)$$

in which the complex Fresnel integral is defined by

$$F(x) = \int_x^\infty \varepsilon^{-j\lambda^2} d\lambda \quad (13)$$

Within the curly bracket the last two terms in each equation correspond to the fields (c) and (d) above, respectively, with the remaining term or terms referring to (b).

It may be noted that, if $\theta_p = \frac{1}{2}\pi$ or $\frac{3}{2}\pi$, the contributions to E_y from the fields (c) and (d) vanish to the order of approximation here considered.

(4.2) The Axial Field

The interaction correction to the axial field is found by integrating around the rim the contribution of each current element. The angle ψ is independent of θ , and the total field is easily determined. Since $I^{int} = \beta I^e$, where I^e is defined by eqn. (6)

of Reference 1, this field resembles closely the first-order axial field found previously when interaction was neglected:

$$E_y(0, 0, z) = \varepsilon^{-jkz} - \frac{\varepsilon^{-jkR}}{2} \left[\sqrt{\left(1 + \frac{a}{R}\right)} (1 + \beta) + \sqrt{\left(1 - \frac{a}{R}\right)} \frac{z}{R} \right] \quad (14)$$

where

$$\beta = -\frac{a}{a+R} \frac{\varepsilon^{-j(2ka - \pi/4)}}{\sqrt{(\pi ka)}} \quad (15)$$

$$R^2 = a^2 + z^2 \quad (16)$$

and

$$E_x(0, 0, z) = E_z(0, 0, z) = 0 \quad (17)$$

The expression (14) is valid if z is not too large, although the interaction correction itself is valid for all values of z .

(5) THE TRANSMISSION COEFFICIENT FOR NORMAL INCIDENCE

The transmission coefficient, t , for an aperture in a screen may be found from knowledge of the far-field in the direction of propagation of the incident plane wave,^{2,11} and for a circular aperture, it is given by

$$t = \lim_{z \rightarrow \infty} \frac{2}{a^2} \mathcal{R} \left[\frac{E_y(0, 0, z)}{jk \frac{\varepsilon^{-jkz}}{z}} \right] \quad (18)$$

For the present discussion, it is necessary to separate E_y into the sum of two components, E_y^0 and E_y' , say. E_y^0 is the incident wave plus the field scattered by the currents flowing in the screen when interaction is neglected. E_y' is the field due to interaction which, on the axis of the aperture, is given by the term dependent on β in eqn. (14). If the corresponding components of t are t^0 and t' , then

$$t = t^0 + t' \quad (19)$$

(5.1) The Transmission Coefficient neglecting Interaction

To determine t^0 , it is now necessary to postulate a current distribution over the entire screen when interaction is neglected.

Let a point on the screen be determined by the polar co-ordinates $(r + a, \theta)$; r is thus measured from the edge of the aperture. The incident electric vector, assumed parallel to the y -axis, makes an angle θ with the tangent to the aperture rim at (a, θ) . Suppose the current density at a point distant r from the edge of a half-plane ($z = 0, X \geq 0$) irradiated by a normally incident plane wave whose electric vector makes an angle θ with the edge of the plane (the positive Y -axis) is $J = (J_x, J_y, 0)$.

Then,

$$\left. \begin{aligned} J_x &= \frac{1}{Z_0} \Psi(kr) \sin 2\theta \\ J_y &= \frac{2}{Z_0} [1 - 2\Phi(kr) - \Psi(kr) \cos^2 \theta] \end{aligned} \right\} \quad (20)$$

where

$$\left. \begin{aligned} \Psi(kr) &= \frac{\varepsilon^{j3\pi/4}}{\sqrt{\pi}} \frac{\varepsilon^{-jkr}}{\sqrt{(kr)}} \\ \Phi(kr) &= \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} F[\sqrt{(kr)}] \end{aligned} \right\} \quad (21)$$

This half-plane current distribution might be chosen as the non-interaction density on the screen for the following reasons: (a) the field near the edge of the (large) aperture is then the half-plane field as required; (b) the current has the correct order of singularity at the edge; (c) the current has the proper behaviour

at very large distances from the aperture. The choice of geometrical-optics current densities $J_x = 0$, $J_y = 2/Z_0$, would violate both (a) and (b).

Employing essentially the half-plane current distribution given above, Frahn¹² has evaluated t^0 and obtained (in error) $t^0 = 1$ for all values of ka . Chang³ obtained $t^0 = 1 - 1/[8(ka)^2]$ with this same distribution, while the present author found $t^0 = 1 - 1/[4(ka)^2]$. This latter value agrees to order $(ka)^{-2}$ with one found by Frahn using an approximation to the tangential electric field in the aperture.

The use of the half-plane currents in the 3-dimensional problem is open to the criticism that, with the exception of the constant component, they fall off as $1/\sqrt{r}$ rather than as $1/r$. This undesirable feature may be eliminated by making use of Chang's solution for the screen current density. By solving approximately the relevant integral equations, Chang obtained the following expressions which will be taken as the screen current density when interaction is neglected:

$$\left. \begin{aligned} J_x &= \frac{1}{Z_0} \sqrt{\left(\frac{a}{r+a}\right)} \Psi(kr) \sin 2\theta \\ J_y &= \frac{2}{Z_0} \left\{ 1 - \sqrt{\left(\frac{a}{r+a}\right)} [2\Phi(kr) + \Psi(kr) \cos^2 \theta] \right\} \end{aligned} \right\} \quad (22)$$

While being very similar to eqns. (20), the expressions (22) for J_x and J_y fall off (with the exception noted above) correctly as $1/r$, and also satisfy the other three conditions previously mentioned.

The axial far-field corresponding to this current density is evaluated in Section 9.2, and gives for t^0 the following expression:

$$t^0 = -\frac{1}{3} + \mathcal{R} \left[\frac{\sqrt{(2\pi\eta)}}{6} e^{-j\pi/4} \varepsilon^{j\eta} \left(1 + 2j \frac{d}{d\eta} - \frac{d^2}{d\eta^2} \right) H_0^{(2)}(\eta) \right. \\ \left. + \frac{\sqrt{\pi}}{4\sqrt{(2\eta)}} \varepsilon^{j\pi/4} \varepsilon^{j\eta} \left(1 + j \frac{d}{d\eta} \right) H_0^{(2)}(\eta) \right] \quad (23)$$

$$\text{where} \quad \eta = \frac{ka}{2} \quad \dots \quad (24)$$

and $H_0^{(2)}(\eta)$ is the zero-order Hankel function of the second kind, defined in terms of the Bessel functions of the first and second kinds by

$$H_0^{(2)}(\eta) = J_0(\eta) - jN_0(\eta) \quad \dots \quad (25)$$

Eqn. (23) may be simplified by carrying through the indicated differentiations and separating real and imaginary parts:

$$t^0 = -\frac{1}{3} + \frac{\sqrt{(\pi\eta)}}{6} \left\{ [A_1(\eta) - A_2(\eta)] \cos \eta + [A_1(\eta) + A_2(\eta)] \sin \eta \right\} \\ + \frac{\sqrt{\pi}}{8\sqrt{\eta}} \left\{ [A_3(\eta) + A_4(\eta)] \cos \eta - [A_3(\eta) - A_4(\eta)] \sin \eta \right\} \quad (26)$$

$$\text{where} \quad \left. \begin{aligned} A_1(\eta) &= \frac{3}{2}J_0(\eta) - 2N_1(\eta) - \frac{1}{2}J_2(\eta) \\ A_2(\eta) &= \frac{3}{2}N_0(\eta) + 2J_1(\eta) - \frac{1}{2}N_2(\eta) \\ A_3(\eta) &= J_0(\eta) - N_1(\eta) \\ A_4(\eta) &= N_0(\eta) + J_1(\eta) \end{aligned} \right\} \quad \dots \quad (27)$$

For large values of ka , t^0 has the following asymptotic expansion:

$$t^0 \simeq 1 - \frac{1}{8(ka)^2} + \frac{45}{512(ka)^4} + 0 \left[\frac{1}{(ka)^6} \right] \quad \dots \quad (28)$$

Since eqn. (26) is non-oscillatory in ka —which may be seen from Fig. 6 or eqn. (28) for large values of ka —it may be safely assumed that the postulated screen current density does, in fact,

contain no interaction component; for interaction from across the aperture must produce terms in t which are oscillatory in ka . [See, e.g., eqns. (14) and (15).]

(5.2) The Interaction Component of the Transmission Coefficient

The interaction component of the transmission coefficient is readily found from the corresponding component of the axial electric field:

$$E'_y(0, 0, z) = -\frac{\beta}{2} \varepsilon^{-jkR} \sqrt{\left(1 + \frac{a}{R}\right)} \quad \dots \quad (29)$$

If $z > a$, E'_y may be expanded in powers of a/z :

$$E'_y(0, 0, z) = \frac{a}{2} \frac{\varepsilon^{-j(2ka - \pi/4)}}{\sqrt{(\pi ka)}} \frac{\varepsilon^{-jkz}}{z} + 0 \left(\frac{1}{z^2} \right) \quad \dots \quad (30)$$

$$\text{Hence} \quad t' = -\frac{\sin(2ka - \pi/4)}{\sqrt{\pi(ka)^{3/2}}} + 0 \left[\frac{1}{(ka)^2} \right] \quad \dots \quad (31)$$

where the terms of order $(ka)^{-2}$ will also be oscillatory in ka , and

$$t = t^0 - \frac{\sin(2ka - \pi/4)}{\sqrt{\pi(ka)^{3/2}}} + 0 \left[\frac{1}{(ka)^2} \right] \quad \dots \quad (32)$$

The asymptotic form of eqn. (32) is, up to and including the term of order $(ka)^{-3/2}$, identical to that found by Chang.

(6) DISCUSSION

Some of the results of the preceding Sections are illustrated in Figs. 4, 5, and 6, and compared with the exact solutions of Andrejewski.¹³⁻¹⁵ Numerical values for these latter have been computed by the present author from the Tables supplied in Reference 13.

Figs. 4(a) and 4(b) exhibit the variation along the magnetic diameter ($y = 0$) of the real and imaginary parts of the total tangential electric field for the particular case $ka = 10$. Near the centre, the approximation in terms of Bessel functions for small values of x/a [eqn. (25) in Reference 1] has been employed, after correcting for interaction through multiplication of the scattered component by $(1 + \alpha)$ [see eqn. (11)]. In the neighbourhood of the rim the approximation given by eqn. (12) is applicable, while in the intermediate region eqn. (4), suitably modified for interaction by the above factor, is valid.

Figs. 4(c) and 4(d) illustrate the field variation along the electric diameter ($x = 0$), using the same methods of evaluation as in Figs. 4(a) and 4(b). Near the rim, as noted in Section 4.1, the field on this diameter is approximately that of an isolated half-plane only, whereas on the magnetic diameter interaction effects play a more important role.

The agreement between approximate and exact theories for this still relatively small aperture is quite gratifying. It appears that, with the aid of these three methods of evaluation, the field at any point in a circular aperture of adequate size may be predicted with good accuracy.

The field at the centre of a circular aperture as a function of ka is illustrated in Fig. 5, where the real and imaginary parts of E_y are taken as abscissa and ordinate, respectively. The curve is obtained from exact theory and the approximation points from eqn. (14), with $z = 0$. Even for $ka \simeq 3$, fairly close agreement between exact and approximate theories may be noted, but not until $ka \simeq 7$ does it become consistently satisfactory. Improvement upon the non-interaction theory for $ka < 10$, in which the points lie on a circle centred at $(1, 0)$ with radius $1/\sqrt{2}$, is most noticeable.

The interaction correction contained in eqn. (14) results in a

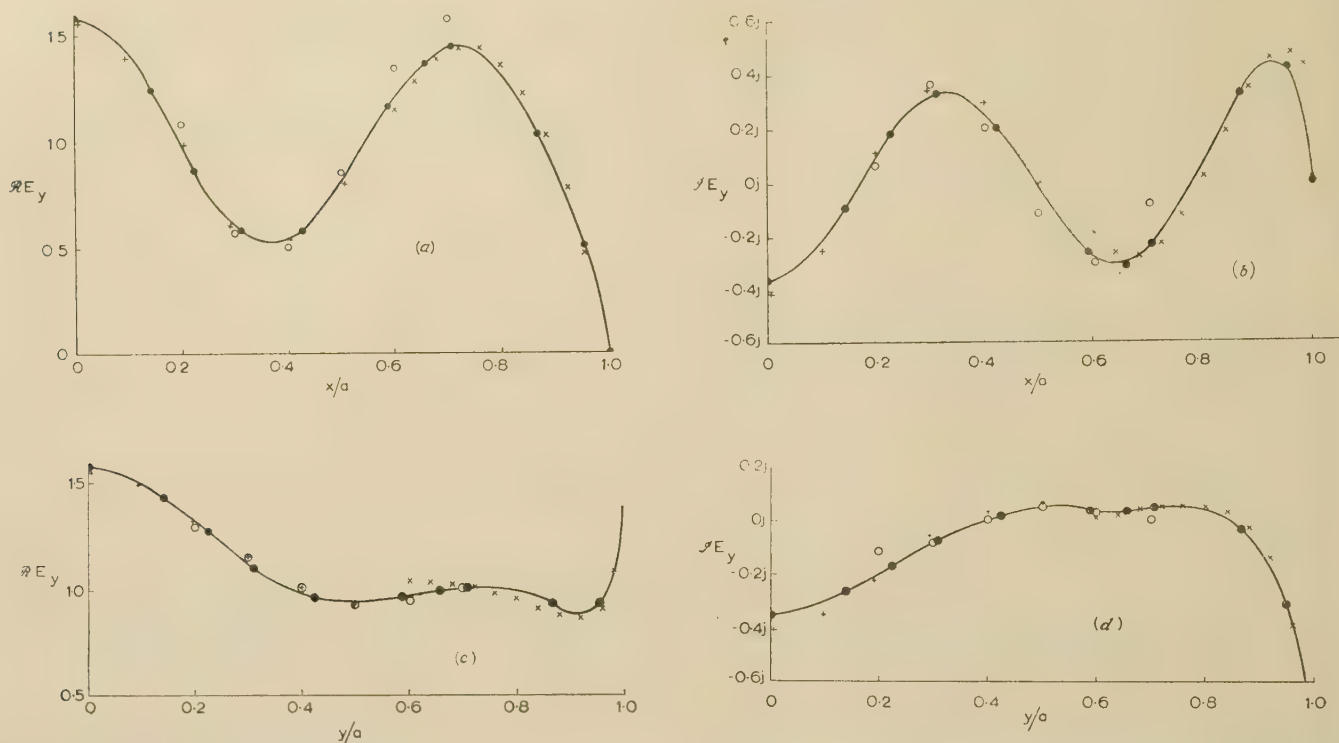


Fig. 4.—Aperture distribution of the electric field along electric and magnetic diameters of a circular aperture; $ka = 10$.

- Exact theory of Andrejewski.
 - + + Edge-current theory: computed from eqns. (24) and (25) in Reference 1 when corrected for interaction.
 - ○ Edge-current theory: computed from eqn. (4) when corrected for interaction.
 - × × Computed from eqn. (12).
- (a) and (b) Along magnetic diameter.
(c) and (d) Along electric diameter.

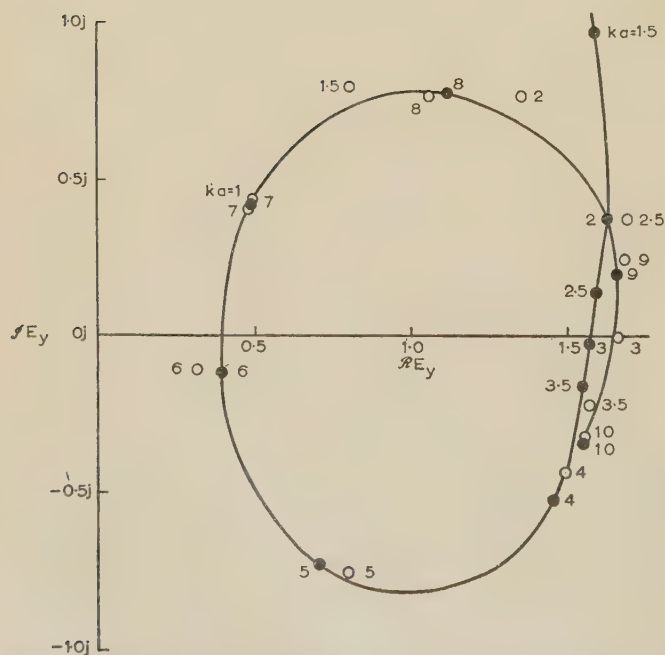


Fig. 5.—Field at the centre of a circular aperture as a function of ka .

- Exact theory of Andrejewski.
- ○ Edge-current theory: computed from eqn. (14).

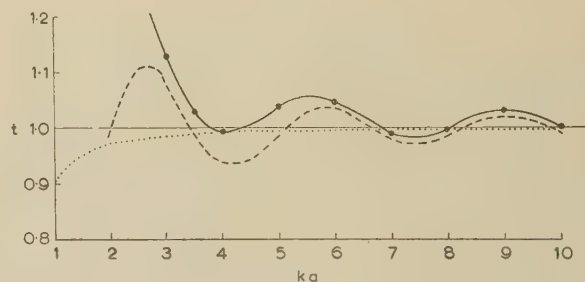


Fig. 6.—Transmission coefficient, t , of a circular aperture for normal incidence.

- Exact theory of Andrejewski.
- - - Edge-current theory: computed from eqn. (32).
- Non-interaction component, t^0 : computed from eqn. (26).

slight decrease in the amplitude of the axial field as compared with the case when interaction is neglected. For $ka \geq 10$, this effect becomes almost negligible, while for smaller apertures the tendency is to narrow the discrepancy between experiment and theory noted in Reference 1.

Fig. 6 shows the variation of the transmission coefficient for normal incidence as a function of ka , as predicted by Andrejewski and by eqn. (32). Calculated from eqn. (26), t^0 is also plotted to demonstrate its monotonic behaviour. Both exact and approximate coefficients appear to exhibit, for $ka \geq 3$, the same oscillatory character, but the approximate values lie consistently below the exact ones. This has been noted also by Chang.

Qualitatively, the exact transmission coefficient seems to possess, for $ka \geq 3$, a monotonically decreasing behaviour in addition to oscillations of form $\sin(2ka - \pi/4)$, whereas the approximate theories predict, as well as these oscillations, a monotonic increase to unity.

A closer fit to rigorous theory may be obtained by determining empirically the desired asymptotic behaviour of t^0 . It appears likely that the first correction term to the unity of geometrical optics is of order $(ka)^{-2}$. In Table 1 the exact values of Andrejewski are compared with $t^0 + t'$, where now $t^0 = 1 + \delta/(ka)^2$, for various values of the parameter δ .

Table 1

COMPARISON BETWEEN EXACT AND SEMI-EMPIRICAL TRANSMISSION COEFFICIENTS

ka	t					
	Exact	$1 + \delta/(ka)^2 + t'$				
		$\delta = 0.60$	0.65	0.70	0.75	0.80
3	1.127	1.161	1.167	1.173	1.178	1.183
4	0.992	0.981	0.983	0.986	0.989	0.993
5	1.039	1.013	1.015	1.017	1.019	1.021
6	1.047	1.054	1.055	1.056	1.058	1.059
7	0.995	0.994	0.995	0.996	0.997	0.998
8	0.999	0.997	0.998	0.999	1.000	1.001
9	1.030	1.028	1.029	1.030	1.030	1.031
10	1.001	1.000	1.001	1.001	1.002	1.002

Although scarcely enough data of a rigorous nature are available to compare with the semi-empirical formula, an examination of Table 1 suggests that δ should lie very close to 0.70 for best agreement in the range $ka \geq 6$.

A monotonic component to the transmission coefficient of a 'soft' screen in the corresponding acoustical problem is not evident, at least to the same extent (see, for example, References 4 and 16). In the electromagnetic case it is probably introduced by the vector nature of the problem.

As mentioned previously, interaction across the aperture would introduce oscillatory terms into the transmission coefficient, and the source of a monotonic term must lie elsewhere. It thus appears that even Chang's approximate solution for the current density on the screen is inadequate to give correctly the monotonic term, although an improvement results on that obtained by employing the unmodified half-plane currents.

The agreement between the first-order interaction component of the transmission coefficient obtained by Chang and by the present method is encouraging, for the simplicity of the basic equations [eqns. (1)] makes it possible to apply the foregoing methods to apertures of more general shape and problems where rigorous theory is intractable, and to obtain a first-order interaction correction to the geometrical-optics transmission coefficient. In addition, this agreement appears to confirm the correctness of the factor $1/\sqrt{2}$ appearing in eqns. (1) as opposed to $1/2$ obtained by Andrews,¹⁷ since the interaction field (and hence the interaction correction to the transmission coefficient) computed from Andrews's equations would contain this discrepancy. (See also Reference 1 for a discussion on this point.)

Additional support is lent to the conclusions of the foregoing analysis by evaluating asymptotically the surface integrals for the aperture field in terms of a current distribution over the entire screen. If the steepest-descents method is applied in turn to the radial and angular integrations, the half-plane currents [eqns. (20)] and Chang's solution [eqns. (22)] both yield the first-order interaction field given by eqns. (5) or (6).

(7) ACKNOWLEDGMENTS

The author would like to thank Dr. P. C. Clemmow for many helpful discussions and useful suggestions and criticisms, Dr. R. D. Kodis for bringing to his attention the work of Chang, and Dr. W. Andrejewski for the loan of a private copy of his dissertation. The financial assistance received from the National Research Council of Canada throughout the period of the research is also gratefully acknowledged.

(8) REFERENCES

- (1) MILLAR, R. F.: 'An Approximate Theory of the Diffraction of an Electromagnetic Wave by an Aperture in a Plane Screen', *Proceedings I.E.E.*, Monograph No. 152 R, October, 1955 (103 C, p. 177).
- (2) LEVINE, H., and SCHWINGER, J.: 'On the Theory of Electromagnetic Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen', *Communications on Pure and Applied Mathematics*, 1950, 3, p. 355.
- (3) CHANG, H. H. C.: 'On the Diffraction of Electromagnetic Waves by a Circular Aperture', Air Force Cambridge Research Center, Cruft Laboratory, Harvard University, Scientific Report No. 2, March, 1955.
- (4) LEVINE, H.: 'Diffraction by a Circular Aperture at High Frequencies', Institute of Mathematical Sciences, Division of Electromagnetic Research, New York University, Research Report No. EM-84, September, 1955.
- (5) KARP, S. N., and RUSSEK, A.: 'Diffraction by a Wide Slit', Institute of Mathematical Sciences, Division of Electromagnetic Research, New York University, Research Report No. EM-75, February, 1955.
- (6) CLEMMOW, P. C.: 'Edge Currents in Diffraction Theory', URSI-Michigan Symposium on Electromagnetic Wave Theory, June, 1955, *Transactions of the Institute of Radio Engineers*, 1956, AP-4, No. 3.
- (7) CARSLAW, H. S.: 'Some Multiform Solutions of the Partial Differential Equations of Physical Mathematics and their Applications', *Proceedings of the London Mathematical Society*, Series 1, 1899, 30, p. 121.
- (8) MACDONALD, H. M.: 'Electric Waves' (University Press, Cambridge, 1902), Appendix D.
- (9) CLEMMOW, P. C.: 'A Note on the Diffraction of a Cylindrical Wave by a Perfectly Conducting Half-Plane', *Quarterly Journal of Mechanics and Applied Mathematics*, 1950, 3, p. 377.
- (10) BAKER, B. B., and COPSON, E. T.: 'The Mathematical Theory of Huygens' Principle' (University Press, Oxford, 1950).
- (11) JONES, D. S.: 'On the Scattering Cross Section of an Obstacle', *Philosophical Magazine*, 1955, 46, p. 957.
- (12) FRAHN, W.: Diplomarbeit, Institut für theoretische Physik, Rheinisch-Westfälische Technische Hochschule, Aachen, 1951.
- (13) ANDREJEWSKI, W.: 'Die Beugung elektromagnetischer Wellen an der leitenden Kreisscheibe und an der kreisförmigen Öffnung im leitenden ebenen Schirm', Dissertation, University of Aachen, 1952.
- (14) ANDREJEWSKI, W.: 'Strenge Theorie der Beugung elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden ebenen Schirm. Numerische Ergebnisse', *Die Naturwissenschaften*, 1951, 38, p. 406.
- (15) ANDREJEWSKI, W.: 'Die Beugung elektromagnetischer Wellen an der leitenden Kreisscheibe und an der kreisförmigen Öffnung im leitenden ebenen Schirm', *Zeitschrift für angewandte Physik*, 1953, 5, p. 178.

- (16) STORRUSTE, A., and WERGELAND, H.: 'On Two Complementary Diffraction Problems', *Physical Review*, 1948, **73**, p. 1397.
- (17) ANDREWS, C. L.: 'Diffraction Pattern in a Circular Aperture measured in the Microwave Region', *Journal of Applied Physics*, 1950, **21**, p. 761.
- (18) JEFFREYS, H., and JEFFREYS, B. S.: 'Methods of Mathematical Physics' (University Press, Cambridge, 1950).

(9) APPENDICES

(9.1) The Asymptotic Evaluation of Eqns. (1)

The substitutions $\theta = \theta_P + \phi$, $\gamma = \theta_P + \chi$, remove the dependence on θ_P from ρ , and result in the following expressions for the scattered field:

$$\left. \begin{aligned} E_x^s &= \frac{1}{2\sqrt{2\pi}} \int_0^{2\pi} \frac{e^{-jk\rho}}{\rho} \cos(\chi - \phi) [\sin(\chi - \phi) + \cos 2\theta_P \sin(\chi + \phi) + \sin 2\theta_P \cos(\chi + \phi)] d\phi \\ E_y^s &= -\frac{1}{2\sqrt{2\pi}} \int_0^{2\pi} \frac{e^{-jk\rho}}{\rho} \cos(\chi - \phi) [\cos(\chi - \phi) + \cos 2\theta_P \cos(\chi + \phi) - \sin 2\theta_P \sin(\chi + \phi)] d\phi \end{aligned} \right\} \quad (33)$$

The stationary points of ρ , given by $d\rho/d\phi = 0$, are readily found to be at $\phi = 0$ and $\phi = \pi$ if $r \neq 0$.

$$\text{Also, } \frac{d^2\rho}{d\phi^2} = \begin{cases} \frac{r}{a-r}, & \phi = 0 \\ -\frac{r}{a+r}, & \phi = \pi \end{cases} \quad \dots \quad (34)$$

It is necessary to consider separately the two stationary points, but for brevity, only that at $\phi = 0$ will be discussed in detail.

The steepest-descent substitutions

$$\rho = \left(1 \mp \frac{r}{a}\right) = \frac{-j}{2} \zeta^2 \quad \dots \quad (35)$$

(where the positive and negative signs are taken accordingly as ϕ is near π or 0, respectively) transform the integrals (33) into the following forms:

$$\left. \begin{aligned} E_x^s &= e^{-jk(a-r)} \int e^{-ka\zeta^2/2} f_1(\zeta) d\zeta \\ &\quad + e^{-jk(a+r)} \int e^{-ka\zeta^2/2} f_2(\zeta) d\zeta \\ E_y^s &= e^{-jk(a-r)} \int e^{-ka\zeta^2/2} g_1(\zeta) d\zeta \\ &\quad + e^{-jk(a+r)} \int e^{-ka\zeta^2/2} g_2(\zeta) d\zeta \end{aligned} \right\} \quad (36)$$

The path of integration is now along the real axis of ζ , in each case passing through $\zeta = 0$.

The functions f_1, f_2, g_1, g_2 may be found as power series in ζ , by employing eqn. (35) and the fact that $\sin \chi = \sin \phi/\rho$:

$$\left. \begin{aligned} f_1(\zeta) &= \sum_{n=0}^{\infty} a_n \zeta^n \\ g_1(\zeta) &= \sum_{n=0}^{\infty} b_n \zeta^n \end{aligned} \right\} \quad \dots \quad (37)$$

with similar expansions for f_2 and g_2 .

The first few coefficients have been found:

$$\left. \begin{aligned} a_0 &= \frac{e^{-j\pi/4}}{2\sqrt{2\pi}} \frac{a}{r^{1/2}} \frac{\sin 2\theta_P}{(a-r)^{1/2}} \\ a_1 &= \frac{-j}{2\sqrt{2\pi}} \frac{a}{r} \left[\frac{r}{a-r} + \left(1 + \frac{a}{a-r}\right) \cos 2\theta_P \right] \\ a_2 &= \frac{e^{j\pi/4}}{4\sqrt{2\pi}} \frac{a}{r^{3/2}} \frac{(4a^2 - 3ar + 2r^2)}{(a-r)^{3/2}} \sin 2\theta_P \\ b_0 &= \frac{e^{-j\pi/4}}{\sqrt{2\pi}} \frac{a}{r^{1/2}} \frac{\cos^2 \theta_P}{(a-r)^{1/2}} \\ b_1 &= -\frac{j}{2\sqrt{2\pi}} \frac{a}{r} \left(1 + \frac{a}{a-r}\right) \sin 2\theta_P \\ b_2 &= -\frac{e^{j\pi/4}}{2\sqrt{2\pi}} \frac{a}{r^{3/2}(a-r)^{3/2}} \{ [a^2 + (a-r)^2] \cos 2\theta_P + ar \cos^2 \theta_P + r^2 \} \end{aligned} \right\} \quad (38)$$

while corresponding expressions may be determined for the coefficients of f_2 and g_2 .

An application of Watson's lemma, which essentially justifies term-by-term integration from $\zeta = -\infty$ to $\zeta = +\infty$ of eqns. (36) (see, for example, chapter 17 of Reference 18), then yields the following asymptotic expansions for the scattered field components in the aperture:

$$\begin{aligned} E_x^s &\approx \frac{\sin 2\theta_P}{2\sqrt{(\pi)r^{1/2}} (ka)^{1/2}} \left\{ \left[\frac{e^{-jk(a-r)-j\pi/4}}{(a-r)^{1/2}} + \frac{e^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \right] \right. \\ &\quad + \frac{1}{2r} \frac{1}{ka} \left[\frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{3/2}} (4a^2 - 3ar + 2r^2) \right. \\ &\quad \left. \left. + \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{3/2}} (4a^2 + 3ar + 2r^2) \right] \right\} + 0 \left[\frac{1}{(ka)^{5/2}} \right] \quad (39) \end{aligned}$$

$$\begin{aligned} E_y^s &\approx -\frac{\cos^2 \theta_P}{\sqrt{(\pi)r^{1/2}} (ka)^{1/2}} \left[\frac{e^{-jk(a-r)-j\pi/4}}{(a-r)^{1/2}} + \frac{e^{-jk(a+r)+j\pi/4}}{(a+r)^{1/2}} \right] \\ &\quad - \frac{\cos 2\theta_P}{2\sqrt{(\pi)r^{3/2}} (ka)^{3/2}} \left[\left(\frac{a-r}{a} + \frac{a}{a-r} \right) \frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{1/2}} \right. \\ &\quad \left. + \left(\frac{a+r}{a} + \frac{a}{a+r} \right) \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{1/2}} \right] \\ &\quad - \frac{1}{2\sqrt{(\pi)r^{1/2}} (ka)^{3/2}} [r + a \cos^2 \theta_P] \\ &\quad \times \left[\frac{e^{-jk(a-r)+j\pi/4}}{(a-r)^{3/2}} + \frac{e^{-jk(a+r)-j\pi/4}}{(a+r)^{3/2}} \right] + 0 \left[\frac{1}{(ka)^{5/2}} \right] \quad (40) \end{aligned}$$

(9.2) The Transmission Coefficient Neglecting Interaction

Eqn. (18) gives an expression for the transmission coefficient in terms of the far-field $E_y(0, 0, z)$. Since, on the axis, the far electric and magnetic fields are related by $E_y(0, 0, z) = -Z_0 H_x(0, 0, z)$ (the total field appearing as a spherical wave radiating from the aperture), an alternative formulation in terms of $H_x(0, 0, z)$ is readily obtained:

$$t = -\lim_{z \rightarrow \infty} \frac{2Z_0}{a^2} \mathcal{D} \left[\frac{H_x(0, 0, z)}{jk \frac{e^{-jkz}}{z}} \right] \quad \dots \quad (41)$$

This expression is more appropriate than eqn. (18) in the

present case, for H_x is given in terms of the y -component of the total Hertz vector, Π_y , only:

$$H_x(x, y, z) = -j\omega\epsilon_0 \frac{\partial \Pi_y(x, y, z)}{\partial z} \quad (42)$$

$$\Pi_y^s(0, 0, z) = \frac{1}{4\pi j\omega\epsilon_0} \iint_{\text{screen}} J_y \frac{e^{-jk\rho}}{\rho} dS \quad (43)$$

where Π_y^s is the scattered component of the Hertz vector, J_y is given by eqns. (22), and

$$\rho^2 = (r+a)^2 + z^2 \quad (44)$$

Adding in the Hertz vector of the incident field,

$$\left[\Pi_y^i(z) = \frac{e^{-jkz}}{k^2} \right]$$

and carrying out some elementary integrations, yields the following expression for the y -component of the total Hertz vector:

$$\Pi_y(0, 0, z) = \frac{ja^2}{2k} \frac{e^{-jkz}}{z} + \frac{2\sqrt{a}}{k} \frac{e^{j3\pi/4}}{\sqrt{\pi}} I_1 - \frac{\sqrt{a}}{2k^{3/2}} \frac{e^{j\pi/4}}{\sqrt{\pi}} I_2 + 0\left(\frac{1}{z^2}\right) \quad (45)$$

$$\text{where } I_1 = \int_0^\infty F[\sqrt{(kr)}] \frac{e^{-jk\rho}}{\rho} \sqrt{(r+a)} dr \quad (46)$$

$$\text{and } I_2 = \int_0^\infty \frac{e^{-jkr}}{\sqrt{r}} \frac{e^{-jk\rho}}{\rho} \sqrt{(r+a)} dr \quad (47)$$

To ensure convergence of the integrals occurring in the subsequent analysis, k will be assumed to possess a small negative imaginary part, which later may be put equal to zero.

If $r+a < z$, ($z > 0$) the following expansion is valid:

$$\frac{e^{-jk\rho}}{\rho} = \frac{e^{-jkz}}{z} + 0\left(\frac{1}{z^2}\right) \quad (48)$$

Hence

$$I_1 = \frac{e^{-jkz}}{z} \int_0^\infty F[\sqrt{(kr)}] \sqrt{(r+a)} dr + \int_{z-a}^\infty F[\sqrt{(kr)}] \left(\frac{e^{-jk\rho}}{\rho} - \frac{e^{-jkz}}{z} \right) \sqrt{(r+a)} dr + 0\left(\frac{1}{z^2}\right) \quad (49)$$

The second integral on the right-hand side may be shown to tend to zero more rapidly than $1/z$ as z approaches infinity, and will therefore be neglected since only terms of order $1/z$ are of interest in the present discussion.

Then, to this order, the following relationship is obtained:

$$I_1 = \frac{e^{-jkz}}{z} \int_0^\infty F[\sqrt{(kr)}] \sqrt{(r+a)} dr \quad (50)$$

and similarly,

$$I_2 = \frac{e^{-jkz}}{z} \int_0^\infty \frac{e^{-jkr}}{\sqrt{r}} \sqrt{(r+a)} dr \quad (51)$$

I_1 may be integrated by parts to give the following:

$$I_1 = \frac{e^{-jkz}}{z} \left[-\frac{a^{3/2}}{3} \sqrt{(\pi)} e^{-j\pi/4} + \frac{\sqrt{k}}{3} \int_0^\infty \frac{e^{-jkr}}{\sqrt{r}} (r+a)^{3/2} dr \right] \quad (52)$$

The substitution

$$r = a \sinh^2 \zeta \quad (53)$$

reduces I_1 and I_2 to the following forms:

$$I_1 = -\frac{e^{-jkz}}{z} \left\{ \frac{a^{3/2}}{3} \sqrt{(\pi)} e^{-j\pi/4} + \frac{\sqrt{k}}{24} \pi a^2 j e^{j\eta} \left[1 + 2j \frac{d}{d\eta} - \frac{d^2}{d\eta^2} \right] H_0^{(2)}(\eta) \right\} \quad (54)$$

$$I_2 = -\frac{e^{-jkz}}{z} \left[\frac{\pi a j}{4} e^{j\eta} \left(1 + j \frac{d}{d\eta} \right) H_0^{(2)}(\eta) \right] \quad (55)$$

where

$$\eta = \frac{ka}{2} \quad (56)$$

and

$$H_0^{(2)}(\eta) = \frac{2j}{\pi} \int_0^\infty e^{-j\eta \cosh \zeta} d\zeta \quad (57)$$

$H_x(0, 0, z)$ is easily found by means of eqns. (42), (45), (54) and (55), from which follows the value of the transmission coefficient exhibited in eqn. (23).

THEORY AND EQUIVALENT CIRCUITS OF THE DOUBLE INDUCTION REGULATOR

By C. S. JHA, B.Sc., Graduate.

(The paper was first received 27th April, and in revised form 19th June, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

Tensor analysis is used to develop a comprehensive theory of the double induction regulator. The general case of a biased double regulator with completely dissimilar component single regulators is first examined. Working equations and equivalent circuits for other types of double regulator are then derived from those of the general case. Tests made on a simple (unbiased) double regulator are shown to agree closely with the behaviour predicted by theory.

(1) INTRODUCTION

For several years, polyphase induction regulators of both single and double types have been in use for voltage control in feeder circuits.¹ The double regulator has recently found widespread use in, and has become an integral part of, the stator-fed polyphase commutator machine. A clear understanding of the theory and characteristics of the double regulator is therefore an essential prelude to the study of this type of machine.

In the polyphase single regulator, which is similar in principle and construction to the normal induction motor working at standstill but with provision for rotating one part (either the primary or the secondary as may be convenient) the secondary electromotive force is constant in magnitude and is regulable only in its phase relation with the primary e.m.f. This limitation of a constant magnitude of the regulable e.m.f. is sometimes undesirable, particularly when the regulator is used for speed control of a.c. commutator machines, and in recent years the double regulators and many modified forms of single regulator have been devised to overcome it.⁴

(2) TYPES OF DOUBLE INDUCTION REGULATOR

Although a number of types of double induction regulator are available to meet specific requirements, there are two main types in common use:

- (a) The simple double regulator.
- (b) The biased double regulator.

(2.1) The Simple Double Regulator

When phase-shift between the secondary and primary e.m.f.'s is not desirable and a variable secondary e.m.f. is required, the simple double regulator is suitable. It consists of two identical single regulators, each of half the total capacity; the primaries are connected in parallel and the secondaries in series, and so arranged that the phase-shift in one regulator neutralizes that in the other, the resultant secondary e.m.f. thus being in phase with the primary e.m.f.

When the shaft of the regulator is turned, the secondary e.m.f. vectors of the two regulators move in opposite directions as shown in Fig. 1(a), and at any instant are equally spaced with respect to the primary e.m.f. vector. Since $|V'_{s1}| = |V'_{s2}|$, the two

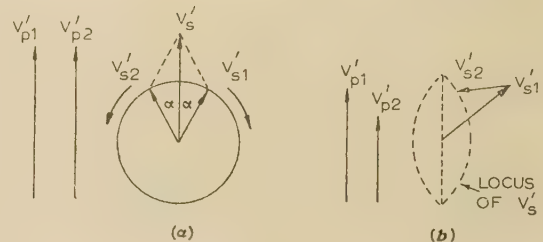


Fig. 1.—Primary and secondary e.m.f. vectors of the double regulator.
(a) Simple regulator. (b) Biased regulator.

regulators being identical, the resultant secondary e.m.f. V'_s is given by

$$V'_s = |V'_{s1}| \cos \alpha + |V'_{s2}| \cos \alpha = 2|V'_{s1}| \cos \alpha$$

where α is the angle by which the secondary e.m.f.'s V'_{s1} and V'_{s2} are displaced in opposite directions from the primary e.m.f. The magnitude of V'_s is thus dependent on the relative displacement between the primary and the secondary windings.

To obtain the effect of opposite directional rotation of the two secondary e.m.f. vectors, two methods are in common use: (a) the rotors of the two regulators are mechanically coupled to rotate in the same direction while the field systems in the two regulators rotate in opposite directions; (b) the rotors are coupled to rotate in opposite directions for any movement of the regulating spindle while the field systems rotate in the same direction.

The currents in the primary windings of the two single regulators will not be in phase, except when both the secondary e.m.f.'s are in phase (or antiphase) or when the secondary is on open-circuit. For this reason the primary windings of all double regulators are connected in parallel and not in series.¹

(2.2) The Biased Double Regulator

In the biased regulator, a small quadrature component is introduced into the secondary e.m.f. by making the secondary e.m.f.'s of the two single regulators unequal. The e.m.f. vectors are shown in Fig. 1(b). The resultant secondary e.m.f. has a locus along the dotted elliptical curve. The quadrature component can be seen to be a maximum when the in-phase component is zero, and zero when the in-phase component is a maximum. The differing secondary e.m.f.'s can be obtained by providing different numbers of turns on the two secondary windings, but as the number of secondary turns is usually small it is generally more convenient to have different numbers of turns on the primaries, thus giving different magnetomotive forces.

(3) THEORY OF THE BIASED DOUBLE REGULATOR

Since the simple double regulator is a special case of the biased double regulator (i.e. in which the two component regulators are exactly alike), the theory of the biased regulator will first be established and that of the simple regulator can be easily derived from it.

Kron's tensor methods,^{3,6} being the best tool available to the modern engineer for the study of the behaviour of electrical

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Jha, formerly at the Heriot-Watt College, Edinburgh, is with the English Electric Co., Ltd.

networks and rotating machines, are used for deriving the equations of performance.

(3.1) Development of Impedance and Admittance Tensors

For the general analysis, the biased double regulator is assumed to have its component single regulators 'a' and 'b' completely dissimilar, i.e. they have different numbers of primary and secondary turns and different effective turns-ratios. In commercial practice, however, either the primaries or the secondaries of the two regulators are alike.

Since the two regulators are each symmetrical on both sides of the air-gap, only single-phase values are used in analysis. The actual and primitive systems are as shown in Fig. 2.

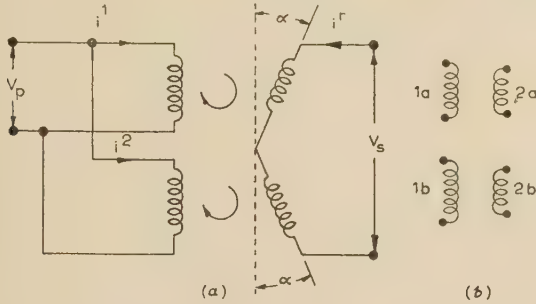


Fig. 2.—Actual (a) and primitive (b) systems for the biased regulator.

The primitive system consists of the four individual coils of the regulator, the two primaries being denoted by '1a' and '1b', and the two secondaries by '2a' and '2b'. The primitive impedance tensor under steady-state conditions is, by inspection, given by

$$Z_{mn} = \begin{matrix} m \backslash n & 1a & 2a & 1b & 2b \\ \begin{matrix} 1a \\ 2a \\ 1b \\ 2b \end{matrix} & \begin{matrix} R_{a1} + jX_{a1} \\ j\omega M_a \\ j\omega M_a \\ R_{b1} + jX_{b1} \end{matrix} & \begin{matrix} j\omega M_a \\ R_{a2} + jX_{a2} \\ j\omega M_b \\ R_{b2} + jX_{b2} \end{matrix} & \begin{matrix} \\ \\ R_{b1} + jX_{b1} \\ j\omega M_b \end{matrix} & \begin{matrix} \\ \\ j\omega M_b \\ R_{b2} + jX_{b2} \end{matrix} \end{matrix} \quad (1)$$

If k_a and k_b be the effective turns-ratios (secondary/primary) of the regulators 'a' and 'b' respectively,

$$\omega M_a = \sqrt{(X_{ma} k_a^2 X_{ma})} = k_a X_{ma} \quad (2)$$

$$\omega M_b = \sqrt{(X_{mb} k_b^2 X_{mb})} = k_b X_{mb} \quad (3)$$

where X_{ma} and X_{mb} are the respective reactances of the primary windings '1a' and '1b' due to mutual flux linking '1a' and '2a' and that linking '1b' and '2b' respectively.

Hence,

$$Z_{mn} = \begin{matrix} m \backslash n & 1a & 2a & 1b & 2b \\ \begin{matrix} 1a \\ 2a \\ 1b \\ 2b \end{matrix} & \begin{matrix} R_{a1} + jX_{a1} \\ jk_a X_{ma} \\ jk_b X_{mb} \\ R_{b1} + jX_{b1} \end{matrix} & \begin{matrix} jk_a X_{ma} \\ R_{a2} + jX_{a2} \\ jk_b X_{mb} \\ R_{b2} + jX_{b2} \end{matrix} & \begin{matrix} \\ \\ R_{b1} + jX_{b1} \\ jk_b X_{mb} \end{matrix} & \begin{matrix} \\ \\ jk_b X_{mb} \\ R_{b2} + jX_{b2} \end{matrix} \end{matrix} \quad (4)$$

For the first transformation from the primitive to the actual system (Fig. 2), the actual system is supposed to have three meshes 1, 2, and r representing the primary of regulator 'a', the primary of regulator 'b' and the secondaries of the two regulators connected in series but displaced from their neutral axes by $+\alpha$ and $-\alpha$ respectively. The connection tensor is established by comparing the actual system currents (i^1 , i^2 , and i^r) with the currents in the primitive system (i^{1a} , i^{2a} , i^{1b} and i^{2b}).

Primitive	Actual
i^{1a}	i^1
i^{2a}	$i^r \epsilon + j\alpha$
i^{1b}	i^2
i^{2b}	$i^r \epsilon - j\alpha$

Since $i^n = C_{\nu}^n i^\nu$; C_{ν}^n is given by

$$C_{\nu}^n = \begin{matrix} n \backslash \nu & 1 & r & 2 \\ \begin{matrix} 1a \\ 2a \\ 1b \\ 2b \end{matrix} & \begin{matrix} 1 \\ \epsilon + j\alpha \\ \\ \epsilon - j\alpha \end{matrix} & \begin{matrix} \\ 1 \\ \\ \end{matrix} & \begin{matrix} \\ \\ 1 \\ \end{matrix} \end{matrix} \quad (5)$$

and the conjugate tensor is C_{μ}^{m*} given by

$$C_{\mu}^{m*} = \begin{matrix} \mu \backslash m & 1a & 2a & 1b & 2b \\ \begin{matrix} 1 \\ r \\ 2 \end{matrix} & \begin{matrix} 1 \\ \epsilon - j\alpha \\ 1 \end{matrix} & \begin{matrix} \\ \epsilon + j\alpha \\ \end{matrix} & \begin{matrix} \\ \epsilon + j\alpha \\ \end{matrix} & \begin{matrix} \\ \end{matrix} \end{matrix} \quad (6)$$

The new impedance tensor $Z_{\mu\nu}$ is given by $Z_{\mu\nu} = C_{\mu}^{m*} Z_{mn} C_{\nu}^n$. Hence,

$$Z_{\mu\nu} = \begin{matrix} \mu \backslash \nu & 1 & r & 2 \\ \begin{matrix} 1 \\ r \\ 2 \end{matrix} & \begin{matrix} R_{a1} + jX_{a1} \\ jk_a X_{ma} \epsilon - j\alpha \\ jk_b X_{mb} \epsilon - j\alpha \end{matrix} & \begin{matrix} jk_a X_{ma} \epsilon + j\alpha \\ (R_{a2} + jX_{a2}) + (R_{b2} + jX_{b2}) \\ jk_b X_{mb} \epsilon + j\alpha \end{matrix} & \begin{matrix} \\ R_{b1} + jX_{b1} \end{matrix} \end{matrix} \quad (7)$$

which reduces to

$$Z_{\mu\nu} = \begin{matrix} \mu \backslash \nu & 1 & r & 2 \\ \begin{matrix} 1 \\ r \\ 2 \end{matrix} & \begin{matrix} Z_{a1} \\ k_a Z_{0a} \epsilon - j\alpha \\ k_b Z_{0b} \epsilon - j\alpha \end{matrix} & \begin{matrix} k_a Z_{0a} \epsilon + j\alpha \\ Z_{a2} + Z_{b2} \\ k_b Z_{0b} \epsilon + j\alpha \end{matrix} & \begin{matrix} \\ Z_{b1} \end{matrix} \end{matrix} \quad (8)$$

where

$$\left. \begin{matrix} Z_{a1} = R_{a1} + jX_{a1} \\ Z_{a2} = R_{a2} + jX_{a2} \\ Z_{b1} = R_{b1} + jX_{b1} \\ Z_{b2} = R_{b2} + jX_{b2} \\ Z_{0a} = jX_{ma} \\ Z_{0b} = jX_{mb} \end{matrix} \right\} \quad (9)$$

Iron losses in the machine can be allowed for by considering Z_{0a} and Z_{0b} to include resistance components.

Since $Y^{\nu\mu} = (Z_{\mu\nu})^{-1}$

$$Y^{\nu\mu} = \begin{array}{c|cc} \nu \backslash \mu & 1 & r & 2 \\ \hline 1 & (Z_{a2} + Z_{b2})Z_{b1} - k_b^2 Z_{0b}^2 & -k_a Z_{0a} Z_{b1} \varepsilon^{+j\alpha} & k_a k_b Z_{0a} Z_{0b} \varepsilon^{+2j\alpha} \\ r & -k_a Z_{0a} Z_{b1} \varepsilon^{-j\alpha} & Z_{a1} Z_{b1} & -k_b Z_{0b} Z_{a1} \varepsilon^{+j\alpha} \\ 2 & k_a k_b Z_{0a} Z_{0b} \varepsilon^{-2j\alpha} & -k_b Z_{0b} Z_{a1} \varepsilon^{-j\alpha} & (Z_{a2} + Z_{b2})Z_{a1} - k_a^2 Z_{0a}^2 \end{array} \times \frac{1}{D} \quad (10)$$

where

$$D = Z_{a1} Z_{a2} Z_{b1} + Z_{a1} Z_{b2} Z_{b1} - K_b^2 Z_{0b}^2 Z_{a1} - K_a^2 Z_{0a}^2 Z_{b1} \quad (11)$$

So far, no transformation has been made to take into consideration the fact that the primaries are connected in parallel. However, if the new voltage tensor is written in the expanded matrix form as

$$v_\mu = \begin{array}{c|c} \mu \backslash r & \\ \hline 1 & V_p \\ & V_s \\ 2 & V_p \end{array} \quad (12)$$

a transformation making the two primaries in parallel is implied (since $v_1 = v_2 = V_p$). Hence the actual performance equations for the regulator can be written down by the relation

$$i^\nu = Y^{\nu\mu} v_\mu \quad (13)$$

Since the two primaries are connected in parallel, the individual primary currents i^1 and i^2 are of little interest from the operational aspect, the actual input current i^p and the output current i^s being of prime importance. The two currents i^p and i^s can be easily calculated from eqn. (13) and the relations $i^p = i^1 + i^2$ and $i^s = i^r$. However, to establish a new impedance tensor between the input and output quantities, a new transformation is useful.

The transformation tensor to operate on $Y^{\nu\mu}$ in eqn. (10) is

written down by comparing the component voltages in the old reference frame (1, r, 2) with those in the new reference frame (p, s), thus

$$\begin{array}{cc} \text{old} & \text{new} \\ V_1 & V_p \\ V_r & V_s \\ V_2 & V_p \end{array}$$

This transformation was inherently implied in eqn. (12). Since $v_\mu = C_{\mu}^{\mu'} v_{\mu'}$

$$C_{\mu}^{\mu'} = \begin{array}{c|cc} \mu \backslash \mu' & p & s \\ \hline 1 & 1 & \\ r & & 1 \\ 2 & 1 & \end{array} \quad (14)$$

The conjugate transformation tensor is obtained by the transpose of $C_{\mu}^{\mu'}$.

Hence,

$$C_{\nu'}^{\nu} = \begin{array}{c|cc} \nu' \backslash \nu & 1 & r & 2 \\ \hline p & 1 & & 1 \\ s & & 1 & \end{array} \quad (15)$$

Since $Y^{\nu'\nu'} = C_{\nu'}^{\nu} Y^{\nu\mu} C_{\mu}^{\mu'}$

$$Y^{\nu'\mu'} = \begin{array}{c|cc} \nu' \backslash \mu' & p & s \\ \hline p & (Z_{a2} + Z_{b2})(Z_{a1} + Z_{b1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha & -(k_a Z_{0a} Z_{b1} \varepsilon^{+j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{-j\alpha}) \\ s & -(k_a Z_{0a} Z_{b1} \varepsilon^{-j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{+j\alpha}) & Z_{a1} Z_{b1} \end{array} \times \frac{1}{D} \quad (16)$$

From the admittance tensor of eqn. (16), the impedance tensor $Z_{\mu'\nu'}$ is obtained by taking the inverse. Hence

$$Z_{\mu'\nu'} = \begin{array}{c|cc} \mu' \backslash \nu' & p & s \\ \hline p & Z_{a1} Z_{b1} & k_a Z_{0a} Z_{b1} \varepsilon^{+j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{-j\alpha} \\ s & k_a Z_{0a} Z_{b1} \varepsilon^{-j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{+j\alpha} & (Z_{a2} + Z_{b2})(Z_{b1} + Z_{a1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha \end{array} \times \frac{1}{DD'} \quad (17)$$

where D' is the value of the determinant

$$|Y^{\nu'\mu'}| \quad (18)$$

From this on simplification,

$$DD' = Z_{a1} + Z_{b1} \quad (19)$$

Hence $Z_{\mu'\nu'}$ reduces to

$$Z_{\mu'\nu'} = \begin{array}{c|c} \begin{array}{c} \mu' \\ \nu' \end{array} & \begin{array}{c} p \\ s \end{array} \end{array} \begin{array}{cc} p & s \\ \hline \frac{Z_{a1}Z_{b1}}{Z_{a1} + Z_{b1}} & \frac{Z_{a1}Z_{b1}}{Z_{a1} + Z_{b1}} \left(k_a \frac{Z_{0a}}{Z_{a1}} e^{+j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} e^{-j\alpha} \right) \\ \hline \frac{Z_{a1}Z_{b1}}{Z_{a1} + Z_{b1}} \left(k_a \frac{Z_{0a}}{Z_{a1}} e^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} e^{+j\alpha} \right) & (Z_{a2} + Z_{b2}) - \frac{k_a^2 Z_{0a}^2 + k_b^2 Z_{0b}^2 - 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha}{Z_{a1} + Z_{b1}} \end{array} \quad (20)$$

(3.2) General Performance Equations

Once the admittance and impedance tensors have been established the equations of performance of the biased double regulator giving relations between input (V_p , i^p) and output (V_s , i^s) quantities can be written down from either of the equations

$$i^{\nu'} = Y^{\nu'\mu'} v_{\mu'}; \text{ and } v_{\mu'} = Z_{\mu'\nu'} i^{\nu'} \quad (21)$$

Here, since

$$i^{\nu'} = \begin{array}{c|c} \begin{array}{c} \nu' \\ p \\ s \end{array} & \begin{array}{c} i^p \\ i^s \end{array} \end{array}; \text{ and } v_{\mu'} = \begin{array}{c|c} \begin{array}{c} \mu' \\ p \\ s \end{array} & \begin{array}{c} V_p \\ V_s \end{array} \end{array} \quad (22)$$

the equations of performance are given by

$$i^p = \frac{1}{D} [(Z_{a2} + Z_{b2})(Z_{b1} + Z_{a1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha] V_p - \frac{1}{D} [k_a Z_{0a} Z_{b1} e^{+j\alpha} + k_b Z_{0b} Z_{a1} e^{-j\alpha}] V_s \quad (a)$$

$$i^s = -\frac{1}{D} (k_a Z_{0a} Z_{b1} e^{-j\alpha} + k_b Z_{0b} Z_{a1} e^{+j\alpha}) V_p + \frac{Z_{a1} Z_{b1}}{D} V_s \quad (b)$$

and

$$V_p = \frac{Z_{a1} Z_{b1}}{Z_{a1} + Z_{b1}} i^p + \frac{Z_{a1} Z_{b1}}{Z_{a1} + Z_{b1}} \times (k_a \frac{Z_{0a}}{Z_{a1}} e^{+j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} e^{-j\alpha}) i^s \quad (c)$$

$$V_s = \frac{Z_{a1} Z_{b1}}{Z_{a1} + Z_{b1}} (k_a \frac{Z_{0a}}{Z_{a1}} e^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} e^{+j\alpha}) i^p + \left[(Z_{a2} + Z_{b2}) - \frac{(k_a^2 Z_{0a}^2 + k_b^2 Z_{0b}^2 - 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha)}{Z_{a1} + Z_{b1}} \right] i^s \quad (d)$$

These four equations are not independent; the last two can be derived from the first two or vice versa, but they have been written down for completeness. They enable the behaviour of the regulator under any steady-state operating conditions to be predicted.

In subsequent analysis the input terminals of the double regulator to which the primaries of the two component single regulators are connected are referred to as *primary* terminals of the double regulator, and the output terminals which comprise

the two secondaries of the component regulators connected in series are referred to as *secondary* terminals of the double regulator. Consequently, the input and output quantities are referred to as primary and secondary quantities of the double regulator. It is to be clearly understood that the primary and secondary quantities of the double regulator do not refer to the corresponding quantities of the component single regulators.

(3.2.1) Open-Circuit and Short-Circuit Conditions.

Since in induction machines the measurement of parameters usually involves open-circuit and short-circuit tests, the behaviour of the double regulator under such conditions is of considerable interest. Four sets of conditions are discussed here.

(a) *Primary terminals of the biased double regulator connected to supply; secondary terminals on open-circuit.*

Under this condition $i^s = 0$. Hence, from (b) of eqn. (23),

$$\frac{V_s}{V_p} = k_a \frac{Z_{0a}}{Z_{a1}} e^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} e^{+j\alpha} \quad (24)$$

and from (c) of eqn. (23),

$$\frac{V_p}{i^p} = \frac{Z_{a1} Z_{b1}}{Z_{a1} + Z_{b1}} \quad (25)$$

From eqn. (24), if Z_{0a}/Z_{a1} and Z_{0b}/Z_{b1} are assumed to be real, the secondary open-circuit voltage V_s has an in-phase component $V_p [k_a (Z_{0a}/Z_{a1}) + k_b (Z_{0b}/Z_{b1})] \cos \alpha$ and a quadrature lagging component $V_p [k_a (Z_{0a}/Z_{a1}) - k_b (Z_{0b}/Z_{b1})] \sin \alpha$. Both components are functions of α , the regulator shaft position. The in-phase component has a maximum numerical value at $\cos \alpha = \pm 1$ (i.e. $\alpha = 0$ or π), where the quadrature component is zero, and a zero value at $\cos \alpha = 0$ (i.e. $\alpha = \pi/2$ or $3\pi/2$), where the quadrature component is a maximum. If the motion

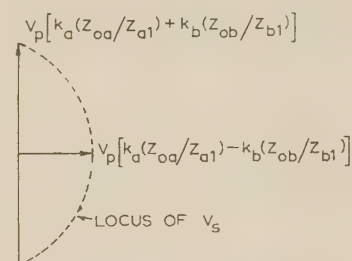


Fig. 3.—Locus of the secondary open-circuit voltage of the biased regulator (α varying between 0 and π).

of the regulator shaft be restricted to allow α to vary only between 0 and π , the locus of the secondary e.m.f. will be as shown in Fig. 3.

Eqn. (25) shows the independence of the input impedance V_p/i^p from the regulator shaft position.

(b) Secondary terminals of the biased double regulator connected to supply; primary terminals on open-circuit.

Under this condition $i^p = 0$. Hence, from (a) of eqns. (23),

$$\frac{V_p}{V_s} = \frac{k_a Z_{0a} Z_{b1} \varepsilon^{+j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{-j\alpha}}{(Z_{a2} + Z_{b2})(Z_{b1} + Z_{a1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha} \quad (26)$$

and from (d) of eqns. (23),

$$\frac{V_s}{i^s} = \frac{(Z_{a2} + Z_{b2})(Z_{a1} + Z_{b1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha}{Z_{a1} + Z_{b1}} \quad (27)$$

Eqns. (26) and (27) show the complex nature of the dependence of the transformation ratio V_p/V_s and the impedance V_s/i^s on α , the regulator shaft position.

(c) Primary terminals of the biased double regulator connected to supply; secondary terminals on short-circuit.

Under this condition $V_s = 0$. Hence, from (a) of eqns. (23),

$$\frac{V_p}{i^p} = \frac{D}{(Z_{a2} + Z_{b2})(Z_{b1} + Z_{a1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha} \quad (28)$$

and from (d) of eqns. (23),

$$\frac{i^s}{i^p} = - \frac{k_a Z_{0a} Z_{b1} \varepsilon^{-j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{+j\alpha}}{(Z_{a2} + Z_{b2})(Z_{b1} + Z_{a1}) - k_a^2 Z_{0a}^2 - k_b^2 Z_{0b}^2 + 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha} \quad (29)$$

Eqn. (28) demonstrates the dependence of the short-circuit impedance V_p/i^p on the regulator shaft position, while eqn. (29) gives the relation between the primary and secondary currents seen clearly to be a complex function of α .

(d) Secondary terminals of the biased double regulator connected to supply, primary terminals on short-circuit.

Under this condition $V_p = 0$. Hence, from (b) of eqns. (23),

$$\frac{V_s}{i^s} = \frac{D}{Z_{a1} Z_{b1}} \quad (30)$$

and from (c) of eqns. (23)

$$\frac{i^p}{i^s} = - \frac{(k_a Z_{0a} Z_{b1} \varepsilon^{+j\alpha} + k_b Z_{0b} Z_{a1} \varepsilon^{-j\alpha})}{Z_{a1} Z_{b1}} \quad (31)$$

Eqn. (30) shows the complete independence of the short-circuit impedance from the regulator shaft position. When the value of D from eqn. (11) is substituted in eqn. (30) we have

$$\frac{V_s}{i^s} = Z_{a2} + Z_{b2} - k_b^2 \frac{Z_{0b}^2}{Z_{b1}} - k_a^2 \frac{Z_{0a}^2}{Z_{a1}} \quad (32)$$

If the self-impedances Z_{a1} , Z_{b1} , Z_{a2} and Z_{b2} are expressed in terms of mutual and leakage impedances,

$$\left. \begin{aligned} Z_{a1} &= z_{a1} + Z_{0a}; & Z_{b1} &= z_{b1} + Z_{0b} \\ Z_{a2} &= z_{a2} + k_a^2 \frac{Z_{0a}^2}{Z_{a1}}; & Z_{b2} &= z_{b2} + k_b^2 \frac{Z_{0b}^2}{Z_{b1}} \end{aligned} \right\} \quad (33)$$

where z_{a1} , z_{b1} , z_{a2} and z_{b2} represent the leakage impedances of the four windings 1a, 1b, 2a and 2b respectively. Eqn. (32) reduces to

$$\frac{V_s}{i^s} = (z_{a2} + z_{b2}) + k_a^2 \left(\frac{Z_{0a}}{Z_{a1}} \right) z_{a1} + k_b^2 \left(\frac{Z_{0b}}{Z_{b1}} \right) z_{b1} \quad (34)$$

which is equal to

(Leakage impedance of the secondaries in series) +
(Leakage impedance of the primary winding of regulator 'a' referred to its secondary side) +
(Leakage impedance of the primary winding of regulator 'b' referred to its secondary side)

or Total leakage reactance of the regulator referred to its secondary side.

Thus the short-circuit impedance with the primary on short-circuit gives the total leakage reactance of the biased regulator referred to the secondary side and is independent of the regulator shaft position. This is an important point, since it is in definite contrast to the short-circuit impedance with the secondary on short-circuit, which has been shown in eqn. (28) to be a complex function of α .

(4) EQUIVALENT CIRCUITS

Although the general behaviour of any machine can be completely determined from the equations of performance, the establishment of an equivalent circuit for the machine is sometimes very useful for numerical solutions.

(4.1) Establishment of Equivalent Circuits from Impedance Tensors

The general impedance tensor of a 4-terminal network is given by

$$Z_{\alpha\beta} = \begin{array}{c|cc} & \beta & 1 & 2 \\ \hline \alpha & & & \\ \hline 1 & Z_{11} & Z_{12} \\ \hline 2 & Z_{21} & Z_{22} \end{array}$$

which also gives the general voltage equations $V_\alpha = Z_{\alpha\beta} i^\beta$ when expanded as

$$\left. \begin{aligned} V_1 &= Z_{11} i^1 + Z_{12} i^2 \\ V_2 &= Z_{21} i^1 + Z_{22} i^2 \end{aligned} \right\} \quad (35)$$

The equivalent circuit to give the same equations of performance is shown in Fig. 4.

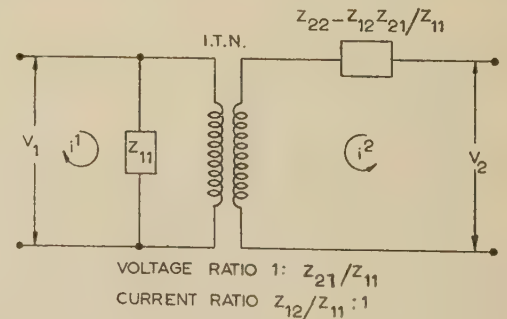


Fig. 4.—Equivalent circuit established from impedance tensors: exact practical type.

From the equivalent circuit may be written

$$V_1 = Z_{11} \left(i^1 + \frac{Z_{12} i^2}{Z_{11}} \right); \quad \text{and} \quad V_1 \frac{Z_{21}}{Z_{11}} = V_2 - i^2 \left(Z_{22} - \frac{Z_{12} Z_{21}}{Z_{11}} \right) \quad (36)$$

which, on simplification, reduce to eqns. (35).

In deriving the equivalent circuit, no assumption as to the interdependence of any of the impedance parameters has been made. However, the concept has been introduced of an ideal transfer network which has specified voltage and current ratios, not necessarily equal. The ideal transfer network can thus be

defined as a hypothetical coupling network with no losses and no leakage reactances, the sole effect of its introduction in any circuit being the transformation in specified ratios of currents and voltages passing through it. Two types of ideal transfer network are in common use in equivalent circuits: (a) the ideal transformer, where the voltage and current transformation ratios are either equal or complex conjugates of each other; and (b) the ideal induction motor, where the voltage ratio is slip times (or the complex conjugate of slip times) the current ratio.

Another type of equivalent circuit which also has the same equations of performance as eqns. (35) is shown in Fig. 5.

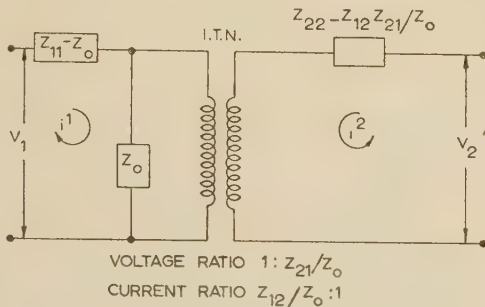


Fig. 5.—Equivalent circuit established from impedance tensors: orthodox type.

This type of equivalent circuit is in common use in the theory of the transformer, in which case $(Z_{11} - Z_0)$ and $[Z_{22} - (Z_{12}Z_{21})/Z_0]$ represent physically the leakage impedances of the two windings 1 and 2, and Z_0 represents the impedance of winding 1 due to the mutual flux between windings 1 and 2.

The equivalent circuits of Figs. 4 and 5 are identical in their equations of performance, but, as shown by Morris⁵ in his exact practical theory of the transformer, the measurement of parameters is much easier for that of Fig. 4. Henceforth, equivalent circuits of the type shown in Fig. 4 will be called exact practical equivalent circuits, while those of the type shown in Fig. 5 will be called orthodox equivalent circuits.

(4.2) Equivalent Circuit of the Biased Regulator

An equivalent circuit for the biased double regulator can be established from its impedance tensor given by eqn. (20) and is shown in Fig. 6, where

$$Z_p = Z_{a1}Z_{b1}/(Z_{a1} + Z_{b1})$$

and

$$Z_l = (Z_{a2} + Z_{b2}) - \frac{(k_a^2 Z_{0a}^2 + k_b^2 Z_{0b}^2 - 2k_a k_b Z_{0a} Z_{0b} \cos 2\alpha)}{Z_{a1} + Z_{b1}}$$

$$- \frac{Z_{a1} Z_{b1} \left(k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{+j\alpha} \right) \left(k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{+j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{-j\alpha} \right)}{Z_{a1} + Z_{b1}}$$

$$= Z_{a2} + Z_{b2} - k_a^2 \left(\frac{Z_{0a}^2}{Z_{a1}} \right) - k_b^2 \left(\frac{Z_{0b}^2}{Z_{b1}} \right) \quad \dots \quad (37)$$

which, from eqns. (33), reduces to

$$Z_l = z_{a2} + z_{b2} + k_a^2 \left(\frac{Z_{0a}}{Z_{a1}} \right) z_{a1} + k_b^2 \left(\frac{Z_{0b}}{Z_{b1}} \right) z_{b1} \quad (38)$$

Z_l in the equivalent circuit thus represents the total leakage impedance of the regulator referred to its secondary side. This equivalent circuit can be seen to be of the exact practical type shown in Fig. 4. Since Z_p , the self-impedance of the two

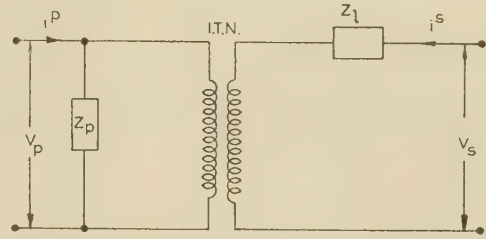


Fig. 6.—Exact practical equivalent circuit of the biased double regulator.

$$\text{Voltage ratio of I.T.N.} = 1 : k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{+j\alpha}$$

$$\text{Current ratio of I.T.N.} = k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{+j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{-j\alpha} : 1$$

primaries in parallel $[= (Z_{a1}Z_{b1})/(Z_{a1} + Z_{b1})]$, cannot be split up into its mutual and leakage components, the establishment of an orthodox equivalent circuit of the type shown in Fig. 5 is not possible.

(4.3) Measurement of Parameters of the Equivalent Circuit

To use the equivalent circuit, knowledge of the following parameters is necessary:

- Z_p , the self-impedance between the primary terminals.
- Z_l , the total leakage impedance referred to the secondary.
- $k_a(Z_{0a}/Z_{a1})\varepsilon^{-j\alpha} + k_b(Z_{0b}/Z_{b1})\varepsilon^{+j\alpha}$, the voltage transformation ratio.
- $k_a(Z_{0a}/Z_{a1})\varepsilon^{+j\alpha} + k_b(Z_{0b}/Z_{b1})\varepsilon^{-j\alpha}$, the current transformation ratio.

From eqn. (25), $Z_p = Z_{a1}Z_{b1}/(Z_{a1} + Z_{b1})$ is given by the input impedance of the primary of the biased regulator when the secondary is on open-circuit, so that a simple open-circuit test is sufficient to determine Z_p . If the power lost in the test is also measured, Z_p can be expressed in the form $R_p + jX_p$, where R_p will include all iron losses. Iron losses, in general, show a tendency to vary slightly with the regulator shaft position, but the effect can be neglected for normal performance calculations. The measurement of Z_p must be carried out at normal operating voltage since its value depends on saturation.

Eqn. (34) shows that Z_l can be measured directly from a short-circuit test with the primary terminals on short-circuit, the value obtained being independent of the shaft position.

It will be seen from eqn. (24) that a simple open-circuit test with the secondary open will give the voltage transformation ratio. This ratio is complex, so that the phase of V_s with respect to V_p must be measured. If V_p is kept constant and the magnitude and phase of V_s for different regulator shaft positions (different values of α) are measured, the value of the voltage transformation ratio for every shaft position can be found. If the shaft positions of the regulator are calibrated in terms of α , it is possible to express the voltage ratio in the more general form $A \cos \alpha + jB \sin \alpha$.

If Z_{0a}/Z_{a1} and Z_{0b}/Z_{b1} can be assumed to be real, the current transformation ratio

$$k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{+j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{-j\alpha}$$

is the complex conjugate of the voltage transformation ratio, the measurement of which has been discussed in the previous paragraph. If more accuracy is desired a current-ratio test with the primary on short-circuit can be made, since from eqn. (31)

$$i_p/i_s = - \left(k_a \frac{Z_{0a}}{Z_{a1}} \varepsilon^{-j\alpha} + k_b \frac{Z_{0b}}{Z_{b1}} \varepsilon^{+j\alpha} \right)$$

This ratio being complex and a function of α has to be, like the voltage transformation ratio, expressed in complex form and measured for all values of $\cos \alpha$.

It is thus seen that an open-circuit test with the secondary open and a short-circuit test with the primary on short-circuit are sufficient to enable measurement of all parameters of the equivalent circuit to be carried out.

(5) BIASED DOUBLE REGULATOR WITH SIMILAR PRIMARIES

When the primaries of the component regulators in the biased double regulator are exactly alike, $Z_{a1} = Z_{b1}$, $Z_{0a} = Z_{0b}$ and $z_{a1} = z_{b1}$. If Z_s , Z_0 and z_s represent the self-impedance, impedance due to mutual flux and leakage impedance respectively of each of the component regulator primaries,

$$Z_s = Z_{a1} = Z_{b1}, Z_0 = Z_{0a} = Z_{0b} \text{ and } z_s = z_{a1} = z_{b1} \quad (39)$$

(5.1) Admittance and Impedance Tensors

From eqns. (39), the admittance and impedance tensors for the general double regulator given by eqns. (16) and (20) respectively reduce to

$$Y^{v'\mu'} = \begin{array}{c} \mu' \backslash v' \\ \begin{array}{cc} p & s \end{array} \\ \begin{array}{cc} p & \begin{array}{c} 2(Z_{a2} + Z_{b2})Z_s - Z_0^2(k_a^2 + k_b^2 - 2k_a k_b \cos 2\alpha) \\ -Z_0 Z_s(k_a e^{+j\alpha} + k_b e^{-j\alpha}) \end{array} \\ s & \begin{array}{c} -Z_0 Z_s(k_a e^{-j\alpha} + k_b e^{+j\alpha}) \\ Z_s^2 \end{array} \end{array} \end{array} \times \frac{1}{D_1} \quad (40)$$

where

$$D_1 = Z_s[(Z_{a2} + Z_{b2})Z_s - (k_a^2 + k_b^2)Z_0^2] \quad (41)$$

and

$$Z_{\mu'\nu'} = \begin{array}{c} \mu' \backslash \nu' \\ \begin{array}{cc} p & s \end{array} \\ \begin{array}{cc} p & \begin{array}{c} \frac{Z_s}{2} \\ \frac{Z_0}{2}(k_a e^{+j\alpha} + k_b e^{-j\alpha}) \end{array} \\ s & \begin{array}{c} \frac{Z_0}{2}(k_a e^{-j\alpha} + k_b e^{+j\alpha}) \\ (Z_{a2} + Z_{b2}) - \frac{Z_0^2}{2Z_s}(k_a^2 + k_b^2 - 2k_a k_b \cos 2\alpha) \end{array} \end{array} \end{array} \quad (42)$$

(5.2) Performance Equations

The equations of performance for the biased regulator with similar primaries are thus given by

$$\left. \begin{aligned} i^p &= \frac{1}{D_1} \left\{ [2(Z_{a2} + Z_{b2})Z_s - Z_0^2(k_a^2 + k_b^2 - 2k_a k_b \cos 2\alpha)] V_p - [Z_0 Z_s(k_a e^{+j\alpha} + k_b e^{-j\alpha})] V_s \right\} \\ i^s &= -\frac{1}{D_1} [Z_0 Z_s(k_a e^{-j\alpha} + k_b e^{+j\alpha}) V_p - Z_s^2 V_s] \end{aligned} \right\} \quad (43)$$

(5.3) Equivalent Circuits

Since the self-impedance of the primary, $Z_s/2$, can be easily split up into its mutual ($Z_0/2$) and leakage ($z_s/2$) components, equivalent circuits of both the practical and orthodox types can be easily established and are shown in Fig. 7.

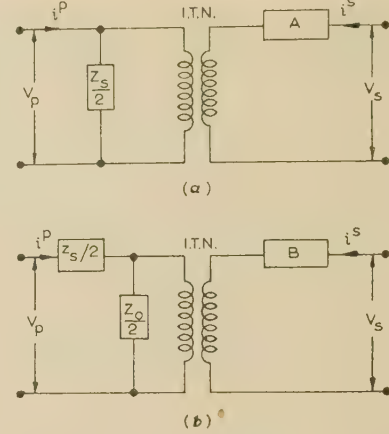


Fig. 7.—Equivalent circuits of biased regulators with similar primaries.

(a) Exact practical circuit.

$$\text{Impedance A} = (z_{a2} + z_{b2}) + (k_a^2 + k_b^2) \frac{Z_0}{Z_s} z_s$$

$$\text{Voltage ratio of I.T.N.} = 1 : \frac{Z_0}{Z_s} (k_a e^{-j\alpha} + k_b e^{+j\alpha})$$

(b) Orthodox circuit.

$$\text{Impedance B} = (z_{a2} + z_{b2}) + \frac{Z_0}{Z_s} (k_a^2 + k_b^2 - 2k_a k_b \cos 2\alpha) \frac{z_s}{2}$$

$$\text{Voltage ratio of I.T.N.} = 1 : k_a e^{-j\alpha} + k_b e^{+j\alpha}$$

(6) BIASED DOUBLE REGULATOR WITH SIMILAR SECONDARIES

When the secondaries of the component single regulators in a biased double regulator are exactly alike, $Z_{a2} = Z_{b2}$ and $z_{a2} = z_{b2}$. If Z_r and z_r represent the self- and leakage-impedances of the secondary of each single regulator,

$$Z_r = Z_{a2} = Z_{b2} \text{ and } z_r = z_{a2} = z_{b2} \quad (44)$$

The impedance and admittance tensors, the equations of performance and the equivalent circuit for this type of biased regulator can be obtained by substituting Z_r for Z_{a2} and Z_{b2} , and z_r for z_{a2} and z_{b2} in the corresponding quantities for the general biased double regulator (Sections 3 and 4).

(7) SPECIAL CASE OF THE SIMPLE DOUBLE REGULATOR

When the two component single regulators are exactly alike, the biased double regulator reduces to a simple double regulator. Hence in a simple double regulator

$$\left. \begin{aligned} Z_{a1} &= Z_{b1} = Z_s; & Z_{a2} &= Z_{b2} = Z_r \\ Z_{0a} &= Z_{0b} = Z_0; & k_a &= k_b = k \\ z_{a1} &= z_{b1} = z_s; & \text{and } z_{a2} &= z_{b2} = z_r \end{aligned} \right\} \quad (45)$$

(7.1) Admittance and Impedance Tensors

Making the substitutions of eqns. (45) in eqns. (16) and (20), the admittance and impedance tensors for the simple double regulator are given by

$$Y^{\nu'\mu'} = \begin{array}{c} \mu' \\ \swarrow \searrow \\ \begin{array}{cc} p & s \\ \hline p & 4Z_r Z_s - 4k^2 Z_0^2 \sin^2 \alpha & -2k Z_0 Z_s \cos \alpha \\ s & -2k Z_0 Z_s \cos \alpha & Z_s^2 \end{array} \end{array} \times \frac{1}{D_2} \quad (46)$$

$$\text{where } D_2 = 2Z_s(Z_r Z_s - k^2 Z_0^2) \quad (47)$$

$$\text{and } Z_{\mu'\nu'} = \begin{array}{c} \mu' \\ \swarrow \searrow \\ \begin{array}{cc} p & s \\ \hline p & \frac{Z_s}{2} & k Z_0 \cos \alpha \\ s & k Z_0 \cos \alpha & 2Z_r - 2\frac{k^2 Z_0^2}{Z_s} \sin^2 \alpha \end{array} \end{array} \quad (48)$$

(7.2) Equations of Performance

From the admittance and impedance tensors the equations of performance are

$$\left. \begin{aligned} i^p &= \frac{4Z_r Z_s - 4k^2 Z_0^2 \sin^2 \alpha}{D_2} V_p - \frac{2k Z_0 Z_s \cos \alpha}{D_2} V_s \quad (a) \\ i^s &= -\frac{2k Z_0 Z_s \cos \alpha}{D_2} V_p + \frac{Z_s^2}{D_2} V_s \quad (b) \\ \text{and } V_p &= \frac{Z_s}{2} i^p + (k Z_0 \cos \alpha) i^s \quad (c) \\ V_s &= (k Z_0 \cos \alpha) i^p + \left(2Z_r - 2\frac{k^2 Z_0^2}{Z_s} \sin^2 \alpha \right) i^s \quad (d) \end{aligned} \right\} \quad (49)$$

(7.3) Equivalent Circuits

Since the impedance tensor is symmetric, a number of equivalent circuits can be easily established.⁷ The orthodox and practical equivalent circuits are shown in Fig. 8.

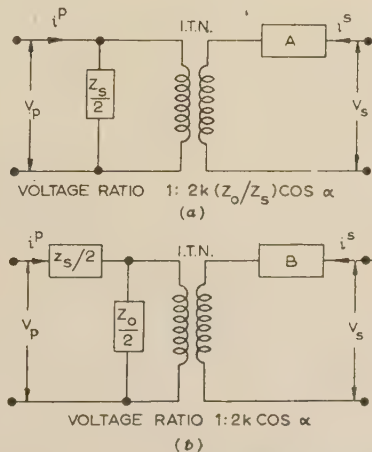


Fig. 8.—Equivalent circuits of simple double regulator.

(a) Exact practical circuit.

$$\text{Impedance A} = 2Z_r + 2k^2 \frac{Z_0}{Z_s} z_s$$

(b) Orthodox circuit.

$$\text{Impedance B} = 2Z_r + 2k^2 \frac{Z_0}{Z_s} z_s \sin^2 \alpha$$

(7.4) Measurement of Parameters

Measurement of parameters of the exact practical equivalent circuit is very simple, since $Z_s/2$ is given by the open-circuit impedance of the regulator with the secondary open, $2k(Z_0/Z_s) \cos \alpha$ is the ratio of the secondary open-circuit voltage to the applied primary voltage at the regulator displacement α , and $2Z_r + 2k^2(Z_0/Z_s)z_s$ is the short-circuit impedance of the regulator with the primary terminals short-circuited. When using the orthodox circuit or the equations of performance it is essential to separate the primary and secondary leakage impedances. If Z_0/Z_s can be assumed to be real, the equivalent resistance in a short-circuit test (primary short-circuited) is

$$R_{eq} = 2R_r + 2k^2 \frac{Z_0}{Z_s} R_s \quad (50)$$

since the equivalent impedance, i.e. the total input impedance in the short-circuit test, is $2Z_r + 2k^2(Z_0/Z_s)z_s$. From the values of the equivalent impedance z_{eq} and its loss component R_{eq} , x_{eq} can be easily measured, since $x_{eq} = \sqrt{(z_{eq}^2 - R_{eq}^2)}$. To separate R_r and R_s , a d.c. test is carried out.

If R and r be the d.c. resistances per phase measured at the input and output terminals of the simple double regulator,

$$R = R_{s(d.c.)}/2 \text{ and } r = 2R_{r(d.c.)} \quad (51)$$

where $R_{s(d.c.)}$ and $R_{r(d.c.)}$ are the values of d.c. resistances of the primary and the secondary of the component single regulators. Hence,

$$R_{eq(d.c.)} = 2R_{r(d.c.)} + 2k^2 \frac{Z_0}{Z_s} R_{s(d.c.)} = r + 4k^2 \frac{Z_0}{Z_s} R \quad (52)$$

This value of $R_{eq(d.c.)}$ should be equal to R_{eq} measured from the power loss in the short-circuit test. Owing to stray losses, however, the value from the short-circuit test is usually larger. If the ratio $R_{eq}/R_{eq(d.c.)}$ be denoted by σ , the respective values of R_s and R_r are given by

$$R_s = 2\sigma R, \text{ and } R_r = \sigma \frac{r}{2} \quad (53)$$

To separate x_{eq} into its component parts x_s and x_r , the assumption, common in dealing with induction machines, is made that the values of the leakage reactances of the primary and secondary referred to the same voltage base are equal. Since the referred value of x_s to the secondary voltage level of the component single regulator is $k^2(Z_0/Z_s)x_s$, which is approximately equal to $k^2|Z_0/Z_s|x_s$, we have

$$\begin{aligned} x_{eq} &= 2x_r + 2k^2 \frac{Z_0}{Z_s} x_s \\ &= 2x_r + 2x_s = 4x_r \quad (54) \end{aligned}$$

Hence, $x_r = x_{eq}/4$ and

$$x_s = \frac{x_r}{k^2} \frac{Z_s}{Z_0} = \frac{x_{eq}}{4k^2} \frac{Z_s}{Z_0} \quad (55)$$

Thus the leakage impedances $z_s (=R_s + jx_s)$ and $z_r (=R_r + jx_r)$ can be determined if k and $|Z_s/Z_0|$ are known. The ratio k is generally known from design figures, while $|Z_s/Z_0|$ can be experimentally measured from an open-circuit test (secondary open) since $2k(Z_0/Z_s) \cos \alpha$ is the ratio of open-circuit secondary voltage to primary applied voltage. If k is not known from design figures, successive approximation should be used to determine exact values for k , z_s and, hence, $Z_s/Z_0 [=Z_s/(Z_s - Z_s)]$, Z_s being twice the input impedance of the regulator with secondary open. Once the value of z_s has been separated the magnetizing impedance $Z_0 (=Z_s - z_s)$ is easily evaluated.

(8) TESTS ON A SIMPLE DOUBLE REGULATOR

A simple double regulator was available for tests. The measurement being based on the theory developed in the previous Sections, the parameters of the test regulator were first measured and the theory was then verified by showing a close agreement between the behaviour of the regulator as calculated from the measured values of parameters and actual test results.

(8.1) The Test Regulator

The 3-phase 50 c/s 2-pole test regulator, supplied as a constituent part of a stator-fed 3-phase shunt commutator motor, has the following rating:

Input: 400 volts, 9.0 amp; Output: 4.13 kVA, 30.6 volts, 45 amp.

(8.2) Measurement of Parameters

Since the shaft movement of the test regulator is not directly graduated in terms of α or $\cos \alpha$, the given arbitrary position marks were calibrated in terms of α as shown in Table 1.

Table 1

CALIBRATION OF REGULATOR-SHAFT POSITION MARKS

Position marks	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	4.8
α , degrees	45	54.37	63.74	73.11	82.48	91.84	101.22	110.59	119.96	129.33	135
$\cos \alpha$	0.707	0.582	0.442	0.291	0.131	-0.033	-0.196	-0.354	-0.5	-0.635	-0.707

The parameters of the test regulator were measured as discussed in Section 7.4, the measured values being as follows:

(a) For the exact practical equivalent circuit of Fig. 8(a):

$$Z_s/2 = 7 + j34.44 \text{ ohms.}$$

$$2kZ_0/Z_s = 2k|Z_0/Z_s| = 0.132.$$

$$2z_r + 2k^2(Z_0/Z_s)z_s = 0.049 + j0.033 \text{ ohm.}$$

(b) For the orthodox equivalent circuit of Fig. 8(b):

$$k = 0.068.$$

$$Z_0 = 12.53 + j67.03 \text{ ohms.}$$

$$z_s = 1.47 + j1.85 \text{ ohms.}$$

$$z_r = 0.0178 + j0.0083 \text{ ohm.}$$

Since, in the equations of performance, all the impedances are in the form of self- or mutual impedances these values of parameters may be written as follows:

(c) For the equations of performance:

$$k = 0.068.$$

$$Z_0 = 12.53 + j67.03 \text{ ohms.}$$

$$Z_s = Z_0 + z_s = 14 + j68.88 \text{ ohms.}$$

$$Z_r = k^2 Z_0 + z_r = 0.0758 + j0.318 \text{ ohm.}$$

In the above measurements, the values for Z_s and Z_0 refer to values obtained at a primary voltage of 400/231 volts. When the magnetic circuit is linear (unsaturated), and the iron losses can be assumed to be proportional to the square of the voltage Z_0 , and Z_s are independent of the voltage applied to the primary terminals. Most iron-cored machines, however, show a magnetic non-linearity due to saturation after a particular value of flux density has been reached.

Fig. 9 shows the input impedance per phase and its loss component between the primary terminals of the test regulator with the secondary terminals open plotted against the applied voltage per phase. The non-linearity can be seen to start when the applied phase voltage reaches approximately 120 volts.

Thus in any actual determination of Z_s (and hence Z_0) the voltage at which the measurements are made is of prime importance. If this voltage differs from the operating voltage, a large error may be introduced in any performance calculations. It can be appreciated that, under any condition of operation, if the applied voltage varies within a very wide range, Z_s and Z_0 no longer remain single-valued unless the range of variation is within the limits of magnetic linearity.

Certain tests and calculations of behaviour have been limited to the linear magnetic range of the regulator. Under this condition of operation the values of Z_s and Z_0 are:

(d) When the magnetic circuit is unsaturated:

$$Z_s = 27 + j105.6 \text{ ohms.}$$

$$Z_0 = Z_s - z_s = 25.5 + j103.75 \text{ ohms.}$$

$$Z_r = z_r + k^2 Z_0 = 0.136 + j0.487 \text{ ohm.}$$

It has been assumed in the above that the variations in the values of leakage impedances, z_s and z_r , under saturated and

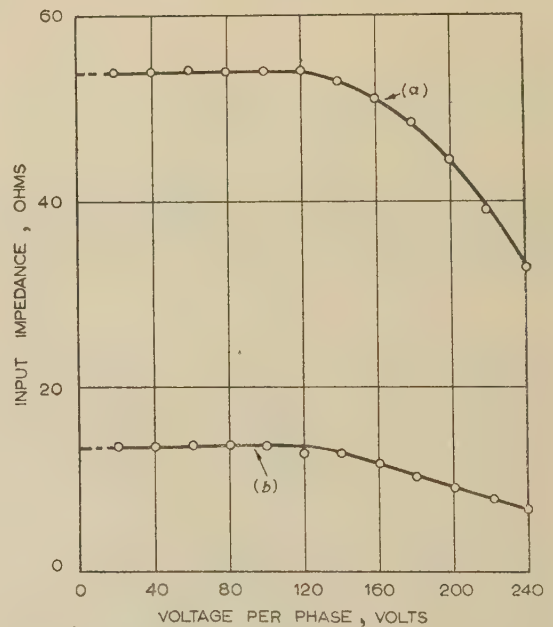


Fig. 9.—Variation of open-circuit input impedance and its loss component with applied primary phase voltage.

(a) Impedance. (b) Loss component of impedance.

unsaturated conditions are negligible. This assumption is justifiable since the leakage impedances are very small compared to the mutual or self-impedances.

(8.3) Calculation of Behaviour and Comparison with Test Results

The equations of performance [eqns. (49)] or either of the two equivalent circuits can be used for calculation of behaviour under

any condition of steady-state operation. The following conditions for the test regulator have been investigated:

- Open-circuit conditions.
- Short-circuit conditions.
- Normal operation of the regulator with a resistive load across the secondary.

(8.3.1) Open-Circuit Conditions: Secondary Open.

Under this condition $i^s = 0$, so that from (c) and (d) of eqns. (49)

$$\left. \begin{aligned} V_s &= (kZ_0 \cos \alpha) i^p \\ V_p &= (Z_s/2) i^p \end{aligned} \right\} (a)$$

and from which

$$\left. \begin{aligned} V_p/i^p &= Z_s/2 \\ V_s/V_p &= 2k(Z_0/Z_s) \cos \alpha \end{aligned} \right\} (b)$$

This shows that the input impedance is independent of the regulator shaft position, while the voltage ratio between the primary and secondary terminals is a linear function of $\cos \alpha$. If the primary voltage be kept constant at 231 volts per phase,

$$\left| \frac{V_p}{i^p} \right| = \left| \frac{Z_s}{2} \right| = |7 + j34.44| = 35.1 \text{ ohms}$$

$$\text{and } V_s = V_p 2k \frac{Z_0}{Z_s} \cos \alpha = 30.5 \cos \alpha \text{ volts}$$

The calculated characteristics and test points for the input impedance and the secondary open-circuit voltage plotted as a function of $\cos \alpha$ are shown in Fig. 10 and show close agreement.

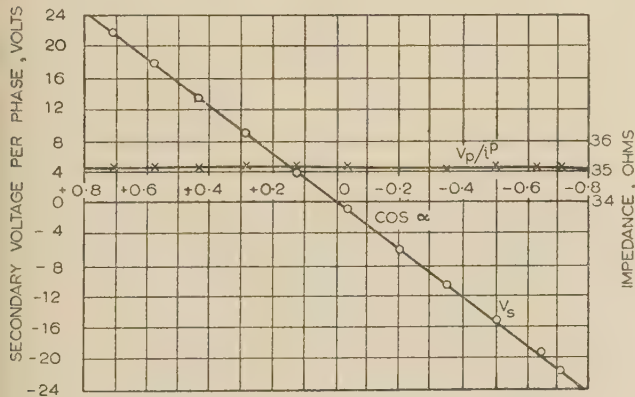


Fig. 10.—Open-circuit condition: secondary open.

Variation of input impedance, V_p/i^p , and secondary open-circuit voltage, V_s , with $\cos \alpha$.

- Calculated curves.
- × Test points for V_p/i^p .
- Test points for V_s .

(8.3.2) Open-Circuit Condition: Primary Open.

Under this condition $i^p = 0$. Hence, from (c) and (d) of eqns. (49),

$$\left. \begin{aligned} V_p &= (kZ_0 \cos \alpha) i^s \\ V_s &= [2Z_r - (2k^2 Z_0^2/Z_s) \sin^2 \alpha] i^s \end{aligned} \right\} (57)$$

If i^s is kept constant, the induced primary voltage V_p is a linear function of $\cos \alpha$ and will vary within the limits 0 at $\cos \alpha = 0$ and $0.707 k Z_0 i^s$ at $\cos \alpha = 0.707$. To keep Z_0 as single-valued the test was carried out with a value of i^s such that the

maximum induced primary voltage did not exceed 120 volts thus keeping the operation within limits of magnetic linearity.

It is interesting to note that the input open-circuit impedance (i.e. V_s/i^s), unlike the case in Section 8.3.1, is a function of the regulator shaft position. Using values of Z_0 , Z_s and Z_r as given in Section 8.2 (d),

$$V_p = 7.28 i^s \cos \alpha \text{ volts}$$

and the open-circuit impedance is

$$\left| \frac{V_s}{i^s} \right| = |(0.049 + j0.033) + (0.0231 + j0.94) \cos^2 \alpha| \text{ ohms}$$

The calculated characteristics and test points for V_p and $|V_s/i^s|$ at $i^s = 25$ amp are shown in Fig. 11 and are in close agreement.

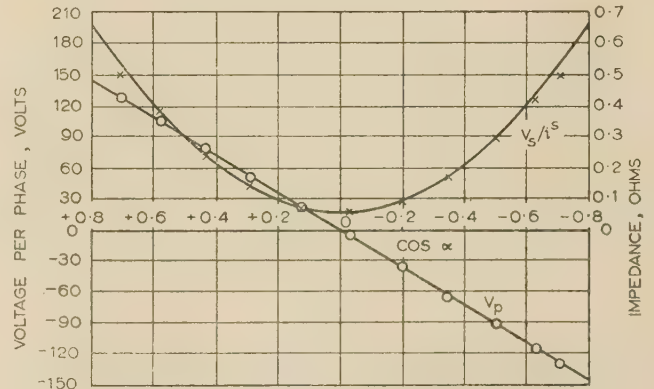


Fig. 11.—Open-circuit condition: primary open.

Variation of input impedance V_s/i^s and open-circuit voltage V_p with $\cos \alpha$, when i^s is kept constant at 25 amp.

- Calculated curves.
- × Test points for V_p/i^s .
- Test points for V_p .

(8.3.3) Short-Circuit Condition: Secondary Short-Circuited.

Under this condition $V_s = 0$, so that from (a) and (d) of eqns. (49) the short-circuit impedance is

$$\left. \begin{aligned} V_p &= \frac{Z_s(Z_r Z_s - k^2 Z_0^2)}{2(Z_r Z_s - k^2 Z_0^2 \sin^2 \alpha)} i^p \\ i^s &= -i^p k Z_0 \cos \alpha / \left(2Z_r - 2k^2 \frac{Z_0^2}{Z_s} \sin^2 \alpha \right) \end{aligned} \right\} (58)$$

and the short-circuit current is

$$i^s = -i^p k Z_0 \cos \alpha / \left(2Z_r - 2k^2 \frac{Z_0^2}{Z_s} \sin^2 \alpha \right)$$

Eqns. (58) show that with the secondary on short-circuit, the short-circuit impedance of the double regulator is a complex function of the shaft position, and is equal to the open-circuit impedance at $\cos \alpha = 0$, at which value the short-circuit current i^s is, of course, zero.

Substituting values of parameters for the test regulator [Section 8.2 (d)] the short-circuit impedance

$$\frac{V_p}{i^p} = 1/[(0.00452 - j0.0177) + (0.247 - j0.167) \cos^2 \alpha] \text{ ohms}$$

and the short-circuit current is

$$i^s = i^p \left[\frac{0.0061 - j0.0056}{\cos \alpha} + (0.132 + j0.0012) \cos \alpha \right]$$

The calculated characteristics and test points for V_p/i^p and for its real part as well as for i^s when $i^p = 3$ amp are shown in

Figs. 12A, 12B and 13. Agreement is fairly close except in the region where $\cos \alpha$ is very small. The discrepancy can be explained to be due to the magnetic circuit becoming saturated in this region as the applied voltage goes beyond the limits of linearity for $i^p = 3$ amp.

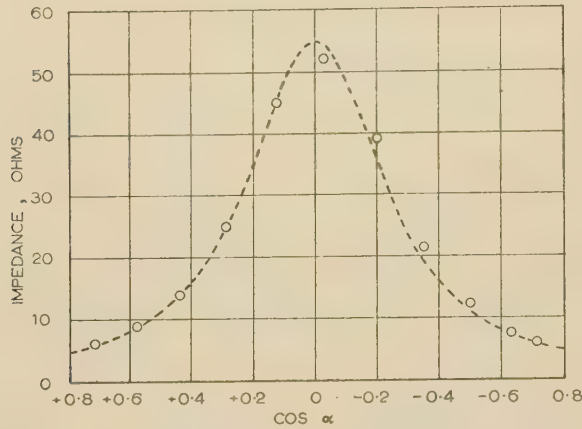


Fig. 12A.—Variation of short-circuit impedance with $\cos \alpha$: secondary on short-circuit.

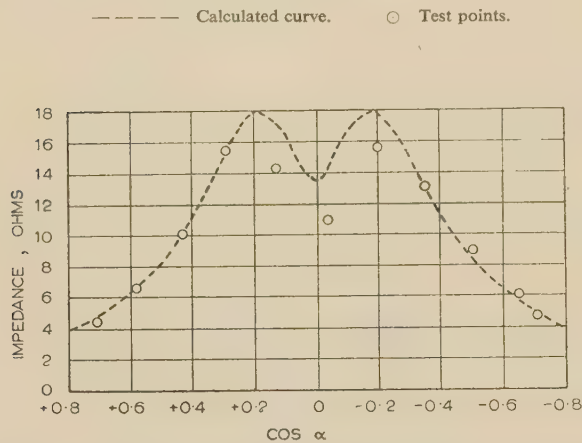


Fig. 12B.—Variation of loss component of short-circuit impedance with $\cos \alpha$: secondary on short-circuit.

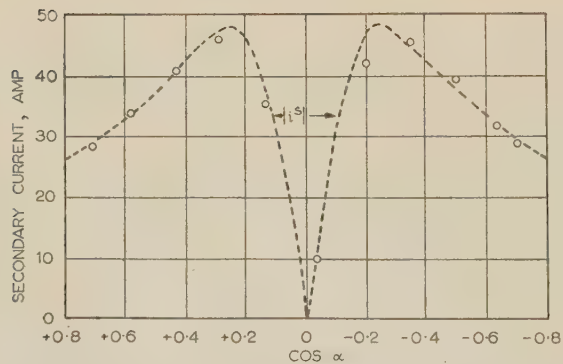


Fig. 13.—Short-circuit condition: secondary on short-circuit.

Variation of secondary current with $\cos \alpha$ when the primary current is kept constant at 3 amp.

— — — Calculated curve. ○ Test points.

(8.3.4) Short-Circuit Condition: Primary Short-Circuited.

Under this condition $V_p = 0$. Hence, from eqns. (49) the short-circuit impedance is

$$\frac{V_s}{i^s} = \frac{2(Z_s Z_0 - k^2 Z_0^2)}{Z_s}$$

and the short-circuit current is

$$i^p = -i^s(2kZ_0/Z_s) \cos \alpha$$

The above equations show that while the short-circuit current is a linear function of $\cos \alpha$, the short-circuit impedance is absolutely independent of the regulator shaft-position. Substituting numerical values, we find the short-circuit impedance to be

$$\left| \frac{V_s}{i^s} \right| = |0.049 + j0.033| \text{ ohm}$$

and the short-circuit current is

$$i^p = i^s(0.132) \cos \alpha$$

The calculated characteristics and test points for the short-circuit impedance $|V_s/i^s|$, for its loss component and for i^p when $i^s = 30$ amp, plotted as functions of $\cos \alpha$, are shown in Fig. 14 and are in close agreement.

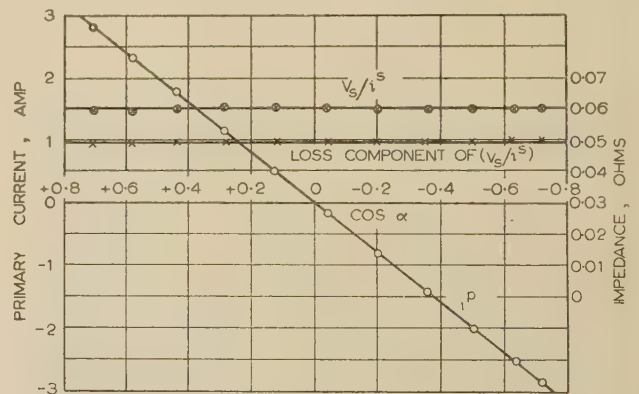


Fig. 14.—Short-circuit condition: primary on short-circuit.

Variation of short-circuit impedance, its loss component and primary current with $\cos \alpha$ when the secondary current is kept constant at 30 amp.

— — — Calculated curves.
⊗ Test points for impedance.
× Test points for loss component of impedance.
○ Test points for primary current.

(8.3.5) Normal Operation of Regulator with a Resistive Load across the Secondary.

A load test was carried out on the regulator with a resistive load across the secondary and the experimental results obtained were compared with those which can be calculated from the equivalent circuits. The operating voltage was 400/231 volts and $\cos \alpha$ was 0.707.

The equivalent circuits for $\cos \alpha = 0.707$ are shown in Fig. 15. It is assumed for simplicity that the load is purely resistive and that the effect of the leakage reactances on the phase of the load current is negligible.

Table 2 gives a comparison of the test values of input current i^p and secondary terminal voltage $|V_s|$ under different load values with those calculated from the two equivalent circuits.

A sample calculation is shown here.

(a) Using the practical equivalent circuit of Fig. 15(a) and a load of 20 amp:

$$i^s = 20 \text{ amp, and } i^{s'} = 0.0935 i^s = 1.87 \text{ amp}$$

$$i^0 = 231/(7 + j34.44) = 1.33 - j6.45 \text{ amp, from which}$$

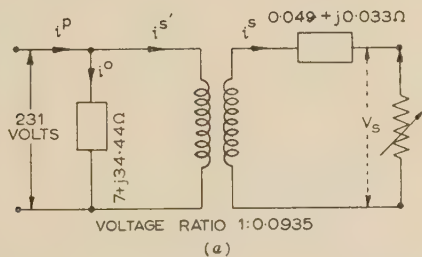
$$i^p = i^0 + i^{s'} = 3.2 - j6.45 \text{ amp}$$

$$V_s = (231 \times 0.0935) - i^s(0.049 + j0.033) \\ = 20.62 - j0.66 = 20.63 \text{ volts.}$$

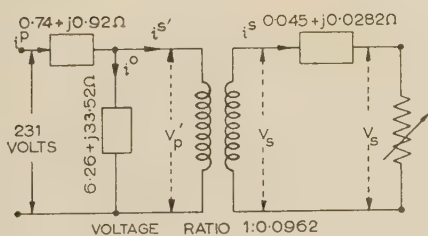
Table 2

COMPARISON OF LOAD TEST RESULTS WITH CALCULATIONS FROM EQUIVALENT CIRCUITS

i^s	i^p			$ V_s $		
	Observed	Calculated from Fig. 15(a)	Calculated from Fig. 15(b)	Observed	Calculated from Fig. 15(a)	Calculated from Fig. 15(b)
amp	amp	amp	amp	volts	volts	volts
0	1.33 $-j6.45$	1.33 $-j6.45$	1.33 $-j6.45$	21.6	21.6	21.6
20	3.18 $-j6.38$	3.2 $-j6.45$	3.2 $-j6.41$	20.5	20.63	20.56
24.6	3.6 $-j6.37$	3.63 $-j6.45$	3.63 $-j6.40$	20.3	20.4	20.32
31.5	4.27 $-j6.36$	4.27 $-j6.45$	4.28 $-j6.39$	19.9	20.06	20.0
37.3	4.84 $-j6.34$	4.82 $-j6.45$	4.83 $-j6.38$	19.6	19.8	19.7



(a)



(b)

Fig. 15.—Equivalent circuits of the test regulator at $\cos \alpha = 0.707$.

(a) Exact practical circuit; (b) Orthodox circuit.

(b) Using the orthodox equivalent circuit of Fig. 15(b) and a similar load:

$$i^s = 20 \text{ amp and } i^{s'} = 0.0962 i^s = 1.924 \text{ amp}$$

$$i^0 = \frac{V_p - i^{s'}(0.74 + j0.92)}{7 + j34.44} = 1.275 - j6.41 \text{ amp}$$

from which

$$i^p = i^0 + i^{s'} = 3.2 - j6.41 \text{ amp}$$

$$V_p' = V_p - i^p(0.74 + j0.92) = 222.74 + j1.8$$

$$V_s = 0.0962 V_p' - i^s(0.045 + j0.0282)$$

$$= 20.55 - j0.39 = 20.56 \text{ volts.}$$

The only approximation made in the practical equivalent circuit is the assumption that Z_0/Z_s can be considered real. Since this is very nearly true, the calculated values of i^p and V_s from the two equivalent circuits are very close to each other. They also agree fairly well with the test results, as can be seen in Table 2. The sample calculations in (a) and (b) above clearly illustrate the advantage of using the practical equivalent circuit for the solution of numerical problems in preference to the orthodox circuit.

(9) CONCLUSIONS

Equations of performance and equivalent circuits for various types of double induction regulators have been established, and the agreement reached between experimental results and results predicted by theory in case of the unbiased double regulator justifies the belief that the theory is basically sound. It is considered that the ease with which the parameters of the exact practical equivalent circuits can be measured and the facility they give in the solution of numerical problems will commend them for wider use by designers and others interested in the theory and operation of double induction regulators.

(10) ACKNOWLEDGMENTS

The investigation was part of a thesis study undertaken by the author, and his thanks are due to the Principal of Heriot-Watt College, Edinburgh, for permission to publish the paper. The author wishes to acknowledge the help and guidance received from Mr. E. O. Taylor, Assistant Professor of Electrical Engineering, Heriot-Watt College, during the investigation and in the subsequent preparation of the paper.

(11) REFERENCES

- (1) CARR, L. H. A.: 'The Use of Induction Regulators in Feeder Circuits', *Journal I.E.E.*, 1925, **63**, p. 864.
- (2) 'The Induction Regulator', *Electrician*, November, 1934, **113**, p. 701.
- (3) KRON, G.: 'A Short Course on Tensor Analysis' (Wiley and Sons, 1942).
- (4) SCHWARZ, B.: 'The Stator-Fed A.C. Commutator Machine with Induction Regulator Control', *Proceedings I.E.E.*, Paper No. 849 U, October, 1949 (**96**, Part II, p. 755).
- (5) MORRIS, D.: 'Some Tests of an Exact Practical Theory of the Transformer', *ibid.*, Paper No. 902 S, February, 1950 (**97**, Part II, p. 17).
- (6) GIBBS, W. J.: 'Tensors in Electrical Machine Theory' (Chapman and Hall, 1952).
- (7) JHA, C. S.: 'Tensor Analysis of the Steady-State Behaviour of the Stator-Fed Polyphase Shunt Commutator Motor'. Thesis for Fellowship of the Heriot-Watt College, June, 1955.

REPRESENTATION OF SALIENCY ON A.C. NETWORK ANALYSERS

By COLIN ADAMSON, M.Sc.(Eng.), Associate Member, and A. M. S. EL-SERAFI, B.Sc., Ph.D.

(The paper was first received 2nd February, and in revised form 10th April, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

Existing methods of representing saliency on network analysers are discussed. These include combined-reactance methods and two methods using a d.c. signal to represent the e.m.f. of the synchronous machine.

A simple method of representing saliency, in association with a network-analyser generator unit, is then described, together with the associated electronic circuits. The results of a problem involving saliency, which has been solved by means of the apparatus described, are presented and the results are discussed.

It is concluded that the method employed is more economical than those previously published, and is well within the normal limits of accuracy for electrical power-system analysis.

LIST OF PRINCIPAL SYMBOLS

- E = Induced e.m.f. in a synchronous machine.
 E_p = Voltage behind Potier reactance.
 E_q = Voltage behind synchronous quadrature-axis reactance.
 E' = Voltage behind transient direct-axis reactance.
 E_q' = Voltage behind transient impedance of a synchronous machine.
 E_r = Voltage behind synchronous direct-axis reactance.
 V_t = Terminal voltage.
 V_R = Voltage behind armature resistance, i.e. $V_R = V_t + IR$.
 $V_S = V_R + \frac{V_{Rd}(X_d - X_q)}{X_q}$, a voltage which, with E and X_d represents the salient-pole synchronous machine.
 V_{Rd} = Component of V_R in the direct-axis.
 I = Load current.
 I_d = Direct-axis component of I .
 I_q = Quadrature-axis component of I .
 I_f = Field current, or excitation proportional to the field current.
 R = Armature resistance.
 X_d = Synchronous direct-axis reactance.
 X_q = Synchronous quadrature-axis reactance.
 X_d' = Transient direct-axis reactance.
 X_p = Potier reactance.
 θ = Angle between vectors E and V_R .
 ϕ = Load power-factor angle.
 ψ = Load angle of the synchronous machine, i.e. angle between vectors E and V_t .

(1) INTRODUCTION

Until recently, the scope of representation of synchronous machines on a.c. impedance-type network analysers was restricted, saliency being either neglected entirely or represented by the artifices described in Section 3 of the paper. There are some important classes of problem involving salient-pole synchronous machines which require the representation of saliency; these include problems involving voltage-regulator action where the changes in terminal voltage depend on saliency, and stability

problems where the output is appreciably affected by saliency when there is a low value of externally connected reactance. Any simple method of representing saliency on a network analyser may also be of importance when the effect of changing the design parameters of synchronous machines is being studied.

It has been claimed, e.g. by Humphrey Davies and Slemon,¹ that the transformer-analogue network analyser possesses advantages over the a.c. impedance-type network analyser in the independent representation of direct and quadrature axes. Recent work has, however, appreciably reduced these advantages; techniques introduced by Van Ness² in the United States, by Bauer³ in Germany and by the authors have greatly extended the facilities of the a.c. impedance-type of analyser in the representation of saliency.

Any scheme which is introduced for the representation of saliency, or which increases in any manner the complexity of network-analyser generator units, must be carefully scrutinized on the grounds of cost. It is generally considered that the minimum number of generators with which a large-scale network analyser should be equipped is twelve, and the authors regard 1.5% of the total cost of such an analyser, commercially built, as an acceptable addition for representing saliency on twelve generator units. The apparatus described in Sections 4 and 5, falls within this limitation.

(2) BASIC VECTOR DIAGRAMS OF THE SALIENT-POLE SYNCHRONOUS MACHINE

(2.1) Steady-state Conditions

The vector relationships of the salient-pole synchronous machine, analysed on the basis of the two-reaction theory, are well understood and have been exhaustively dealt with by a number of authors. Treatment of the vector diagram may be found, for example, in text-books by Concordia,⁴ and Mortlock and Humphrey Davies,⁵ and in a recent paper by Adkins.⁶ Since it is fundamental to the apparatus described below, and since the symbols vary somewhat in the different publications, the diagram is considered briefly at this stage in a simplified form.

The diagram is shown in Fig. 1(a). Since the only field winding of the machine has its axis coincident with the pole axis (the direct axis), it follows that there is no excitation in the quadrature axis. Conventionally, the induced e.m.f., E , in the machine is displaced 90° in time-phase from the excitation, represented by the current I_f . Subtracted from E are the direct- and quadrature-axis synchronous-reactance voltage drops $I_d X_d$ and $I_q X_q$, and the resistance voltage drop IR , to give the terminal voltage V_t .

If the quantities $I_d X_q$ and $I_q X_q$ are added to V_R , the voltage, E_q , behind the quadrature-axis synchronous reactance is obtained. It will be seen that E and E_q are in the same phase.

In practice, the angle ψ between V_t and E is not known and it is more convenient to construct the vector diagram in accordance with Fig. 1(b). In this version the reactance voltage $I X_q$ is known, and may be extended by the addition of $I(X_d - X_q)$; then by drawing a line, perpendicular to E_q , through the extremity of this latter vector, E is obtained.

Correspondence on Monographs is invited for consideration with a view to publication:

Dr. El-Serafi is at the University of Manchester on leave from Cairo University.
 Mr. Adamson is in the Power Systems Laboratory, Faculty of Technology, University of Manchester.

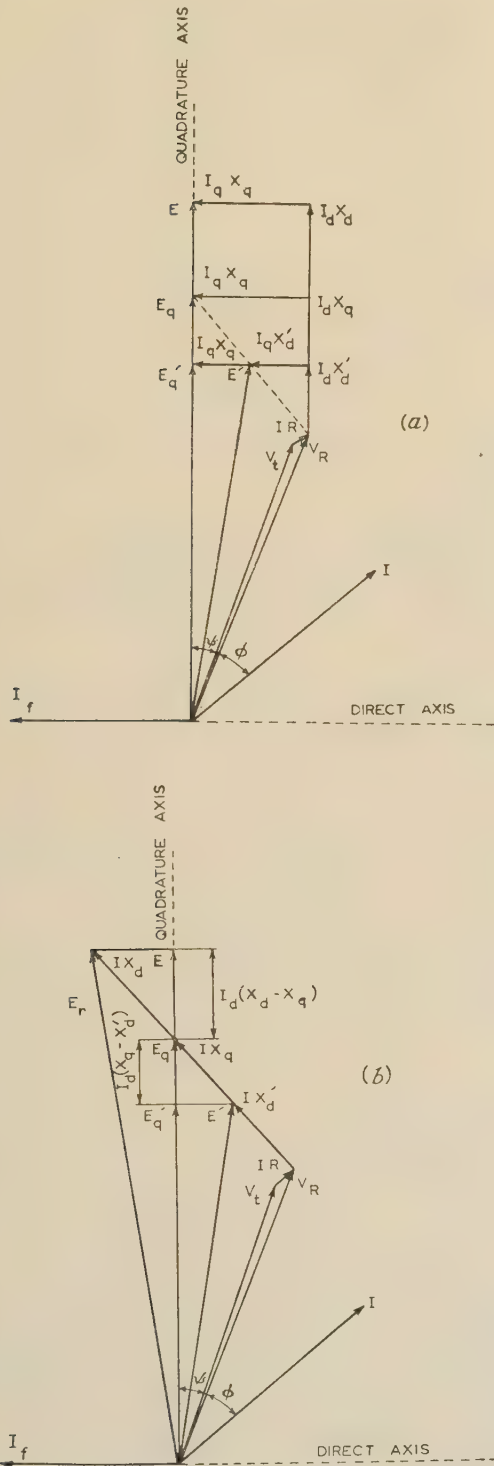


Fig. 1.—Vector diagrams of the salient-pole synchronous machine.

(a) Conventional form.
(b) Modified form.

(2.2) Transient Conditions following an External Disturbance

When problems involving sudden disturbances in the externally connected circuit are considered, and when the effect of damper windings is negligible, the transient direct-axis reactance and synchronous quadrature-axis reactance are used. Events at such instants as this may be represented on the alternator vector

diagram as shown in Figs. 1(a) and 1(b), the problem being reduced to a steady-state basis. Thus, in accordance with the theorem of constant flux linkages,⁷ immediately subsequent to the disturbance a new voltage E'_q is defined, this being the voltage behind the transient reactance. The value of E'_q , together with the direct- and quadrature-axis transient reactances, represents the performance of the machine at the instant under consideration.

(3) EXISTING METHODS OF REPRESENTING SYNCHRONOUS MACHINES ON NETWORK ANALYSERS

(3.1) Methods neglecting Saliency

Many problems in electrical power-system analysis neglect saliency, the synchronous machine being represented by a voltage behind a reactance. For steady-state problems, such as load flow and reactive-power distribution in a network, the reactance adopted is the synchronous reactance, and for balanced-fault and unbalanced-fault studies the reactance is usually the direct-axis transient reactance X'_d ; this procedure is usually adopted on the grounds of simplicity. For the case of a single synchronous machine supplying a network, the use of X'_d is easy to justify on account of the low power factor of the faulted network; as far as the authors are aware, little attention has, however, been paid to the problem of a network supplied by a number of machines, the angles made by their individual direct-axes, with respect to a common reference vector, being different in each case.

In steady-state problems such as voltage regulation and the effect of saturation, it is practically convenient to use the concept of 'Potier reactance', X_p , which is discussed in a number of reference books, including that of Wagner.⁸ When X_p is used on the network analyser, it is combined with X_d , in such a way that two separate reactances are set up in series; these are $(X_d - X_p)$ and X_p , the voltage behind $(X_d - X_p)$ being E and that behind X_p being E_p .

(3.2) Combined-Reactance Methods allowing for Saliency

It is impossible to represent the reactances for both direct and quadrature axes on a network analyser with a single reactance. Saliency may be represented, however, by combined-reactance methods, although these involve trial-and-error procedures.

(3.2.1) Use of E_q .

One method, described by Crary⁹ and Kron¹⁰, represents the synchronous machine by the voltage behind the synchronous quadrature-axis reactance, E_q , in series with two reactances, as shown in Fig. 2(a), and with reference to the vector diagram of Fig. 1(b). The transient direct-axis reactance X'_d , and the difference between the synchronous quadrature-axis reactance and the transient direct-axis reactance $(X_q - X'_d)$ are the quantities involved. The voltage behind X'_d will be E'_q ; this voltage is obtained by neglecting transient saliency. The voltage E_q is not known, but either the quadrature-axis component of E' , i.e. E'_q , or E is known, and E_q must be found by trial and error bearing in mind that

$$E_q + I_d(X_d - X_q) = E = \text{constant} \quad (\text{for steady-state conditions}) \quad (1)$$

$$\text{or } E_q - I_d(X_q - X'_d) = E'_q = \text{constant} \quad (\text{for transient conditions}) \quad (2)$$

(3.2.2) Use of E_r .

Another method, due to Kron,¹⁰ represents the synchronous machine by the voltage behind the direct-axis reactance E_r , as in Fig. 1(b), in series with three reactances, as shown in Fig. 2(b); these reactances are X'_d , $(X_q - X'_d)$, and $(X_d - X_q)$, and

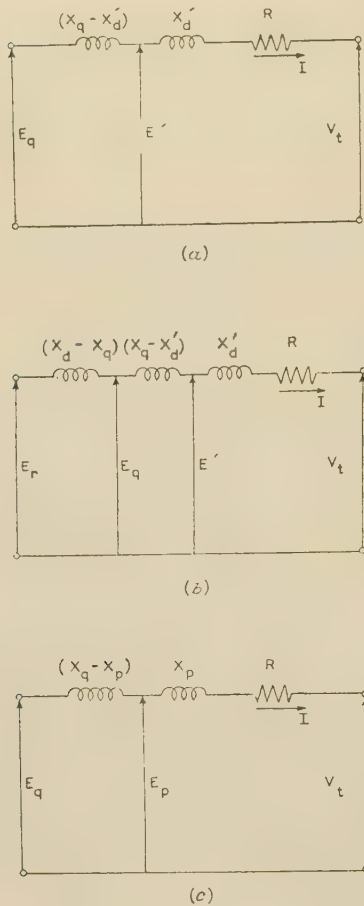


Fig. 2.—Combined-reactance methods of representing saliency.

(a) and (b) Circuits described by Crary and Kron.
 (c) Potier reactance method.

the voltages, E' and E_q , may be measured on the network analyser. In the opinion of the authors this method is better suited to steady-state problems, since E may be obtained directly rather than by calculation from the vector diagram, as in the method discussed in Section 3.2.1. Difficulties do arise in that E_q must be maintained unchanged in phase simultaneously with the adjustment of E_r , in order to satisfy eqns. (1) and (2).

(3.2.3) Use of Potier Reactance.

Fig. 2(c) shows the equivalent circuit for network-analyser representation. The reactance representing the machine, X_q , is split into two components X_p and $(X_q - X_p)$; E_q is the voltage behind the latter reactance, and E_p is the voltage behind the Potier reactance. Generally this concept is used for the problems mentioned in Section 3.1; it is then usual for E to be obtained from E_p by arithmetical methods.

(3.3) Methods of Van Ness and Bauer

(3.3.1) Simulator of Van Ness.

Fig. 3A is a simplified diagram showing the principle of the saliency simulator due to Van Ness.² A d.c. signal proportional to I_d is produced from the generator output and multiplied by $(X_d - X_q)$ or $(X_q - X'_d)$; it is then combined with a d.c. signal proportional to E or E'_q in accordance with eqn. (1) or eqn. (2), respectively. The output is direct current proportional to E_q and must be applied to a modulator for conversion to alternating

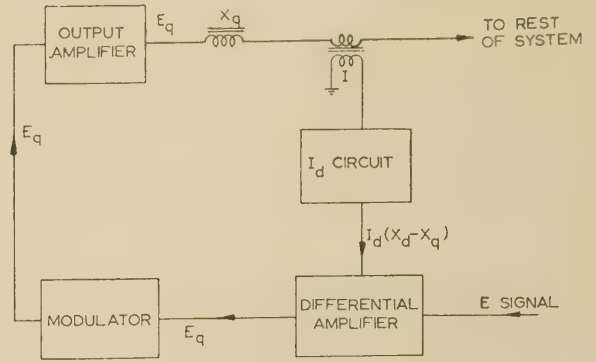


Fig. 3A.—Van Ness simulator.

current, of the appropriate phase, before being fed back to the analyser.

(3.3.2) Simulator of Bauer.

A simplified diagram of the Bauer simulator³ is shown in Fig. 3B. In this case, servo-mechanism methods have been used to generate I_d . A voltage proportional to the current through

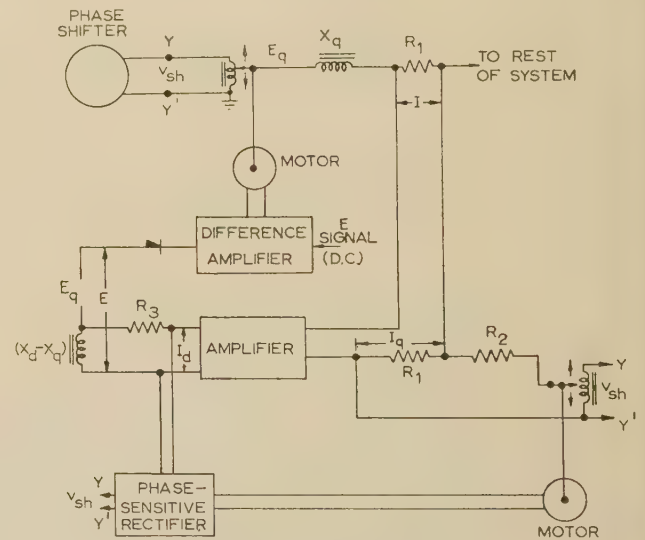
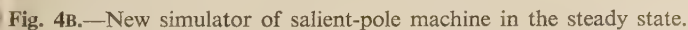


Fig. 3B.—Bauer simulator.

R_1 is subtracted from a controlled voltage in phase with E_q . This gives a voltage proportional to I_d and displaced 90° from E_q , provided that the relation between the controlled voltage and the voltage proportional to current is the same as the relation between I_q and I . Any deviation from the 90° phase displacement between I_d and E_q is fed back to the motor controlling the voltage proportional to I_q in such a manner as to restore the 90° relationship. The voltage proportional to I_d is fed through a high resistance to either $(X_d - X_q)$ or $(X_q - X'_d)$ to give a voltage drop in phase with E_q . Adding or subtracting this voltage from E_q will give E or E'_q . After rectification, the resultant voltage is compared with a direct controlled voltage proportional to E or E'_q ; the difference is fed to a motor to adjust E_q in such a manner that E_q corresponds to either eqn. (1) or eqn. (2).

This saliency simulator is important, since it appears to be the first example of a large-scale network analyser with all its generator units so equipped.



Figs. 4A and 4B show a modified version of the vector diagram of Fig. 1(b) together with a block diagram of the new simulator. A fundamental-frequency voltage, which is always 90° out of phase with the e.m.f. E , and equal to the component of V_R in the direct axis multiplied by a saliency factor $(X_d - X_q)/X_q$, is derived. This voltage is added to V_R , the voltage behind the armature resistance, to give a new voltage V_S . This voltage, together with the e.m.f. E and the direct-axis synchronous reactance, now represents the salient-pole synchronous machine. In the transient condition is to be solved, the voltages behind the subtransient or transient impedances, the corresponding subtransient or transient reactances, are used instead of E , X_d and X_q . It should be noted that the saliency factor may be negative, i.e. the voltage to be added to V_R to give V_S has to be 180° out of phase with that used for the steady-state condition. Referring to the block diagram in Fig. 4B, the scheme uses a

It should be noted that this scheme is automatic; when the working conditions of the machine are changed, the simulator adjusts V_e automatically to represent saliency.

(5.1) Phase-sensitive Rectifier

Fig. 5.—Phase-sensitive rectifier circuit.

rectifier is through a cathode-follower with a transformer output; the secondary winding of this transformer provides a path for the d.c. component of the modulated wave. The switching signal is of large amplitude compared with V_R and is derived very conveniently from the rotor of the magstrip controlling the phase of the input signal to the output amplifier. This magstrip has two rotor windings at right angles; the signal from one of these controls the phase of E and the other is used to feed the square-wave generator with a voltage $E/\pi/2$. The 200-ohm potentiometer on the output is set initially to compensate for unbalance in the rectifier modulator.

(5.2) Square-wave Generator

The square-wave generator is shown in Fig. 6. The signal $E/\pi/2$, which has to be transformed into a square wave, is 'chopped' by two germanium diodes. An equal-interval square wave is obtained by varying the voltage of the common junction of these diodes by a potentiometer chain across the h.t. supply. Direct coupling is used throughout the square-wave generator to reduce differentiation of the square wave. At the output of the final stage, phase splitting takes place, to provide the input to the phase-sensitive rectifier.

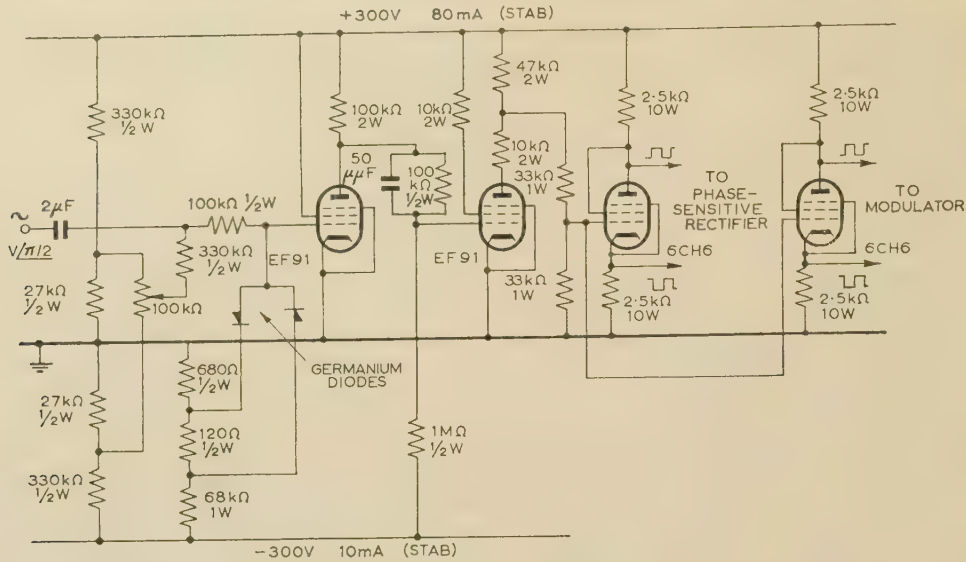


Fig. 6.—Square-wave generator circuit.

(5.3) Modulator

The output of the phase-sensitive rectifier is passed through a low-pass filter so that the d.c. output only remains. The modulator is of the same design as the phase-sensitive rectifier and serves to produce a square wave, the amplitude of which is that of the direct voltage from which it is derived; the phase of this square wave is determined by, and is of the same phase as, that of the square-wave generator. The fundamental component of the square-wave output from the modulator is proportional to $V_R \sin \theta = V_{Rd}$, which is always 90° displaced in phase from E .

To extract this fundamental component a band-pass filter is employed.

(5.4) Introduction of Saliency Factor

The fundamental-frequency voltage V_{Rd} is applied, through an isolating amplifier, to a potentiometer multiplier. This multiplier derives the product $V_{Rd}(X_d - X_q)/X_q$, which is then injected between the terminals of the reactance and the machine resistance, as shown in Fig. 4b; sign-changing is effected by a switch on the output of the injection amplifier.

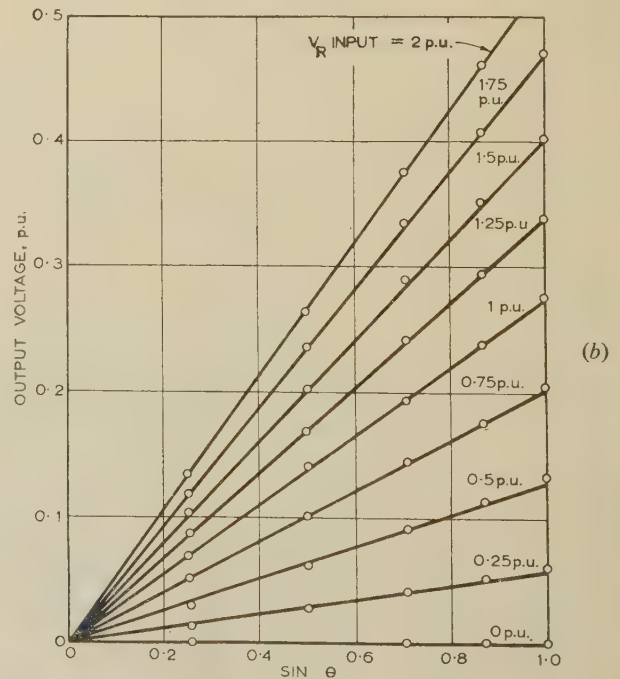
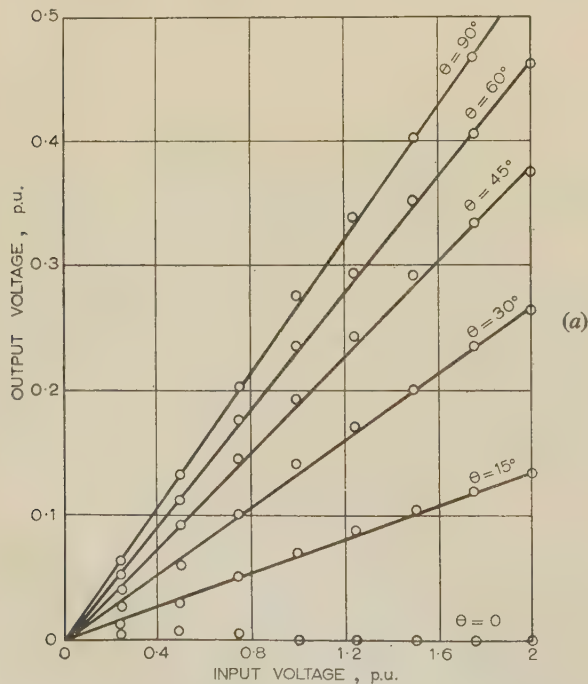


Fig. 7.—Performance of feedback loop of saliency simulator.

(a) Output/input characteristics.

(b) Output/sin θ characteristics.

(6) CHARACTERISTICS OF THE APPARATUS AND RESULTS OF A TYPICAL PROBLEM

(6.1) Characteristics

The overall performance is illustrated by the characteristics of the feedback loop of the complete apparatus, i.e. by the graphs of $(V_S - V_R)$ against V_R for various phase shifts [Fig. 7(a)], and of $(V_S - V_R)$ against $\sin \theta$ for various input voltages V_R [Fig. 7(b)]. These graphs show that the loop is linear within $\pm 3\%$ over the working range.

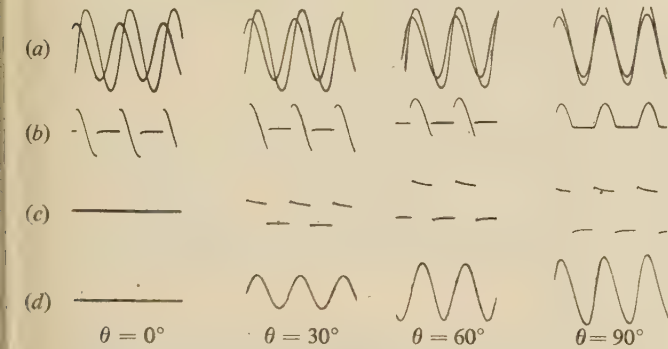


Fig. 8.—Oscillograms at different points on feedback loop.

- (a) Input and signal $E/\sin \theta$
 (b) Output from phase-sensitive rectifier.
 (c) Output from modulator.
 (d) Output from feedback loop.

Fig. 8 shows oscillograms, at various points on the feedback loop, for a series of values of phase shift with $V_R = 1$ p.u. and a saliency factor $(X_d - X_q)/X_q = 0.274$.

(6.2) Study of a Salient-Pole Synchronous-Machine Problem

As a further check on the performance of the simulator, a simple problem, which could also be calculated and in which the effect of saliency was significant, was considered. The problem involved a synchronous generator connected through a reactance to an infinite busbar of voltage 1 p.u., the resistance of the machine being neglected. The constants adopted were as follows:

$$\left. \begin{aligned} X_d &= 1 \text{ p.u.} \\ X_q &= 0.7 \text{ p.u.} \\ X_d' &= 0.4 \text{ p.u.} \end{aligned} \right\} \text{ for the generator.}$$

$$X_e = 0.5 \text{ p.u. for the transmission line.}$$

Four cases were considered:

(a) Steady-state conditions with saliency neglected, i.e. $X_d = X_q = 1$ p.u.

$$\text{Saliency factor} = \frac{(X_d - X_q)}{X_q} = 0.$$

(b) Steady-state conditions with saliency considered, i.e.

$$\frac{(X_d - X_q)}{X_q} = 0.429$$

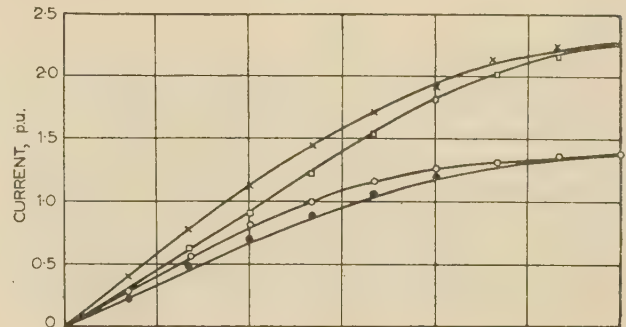
(c) Transient conditions, the problem being solved on a steady-state basis, but with saliency neglected, i.e.

$$X_d' = X_q = 0.4 \text{ p.u.}$$

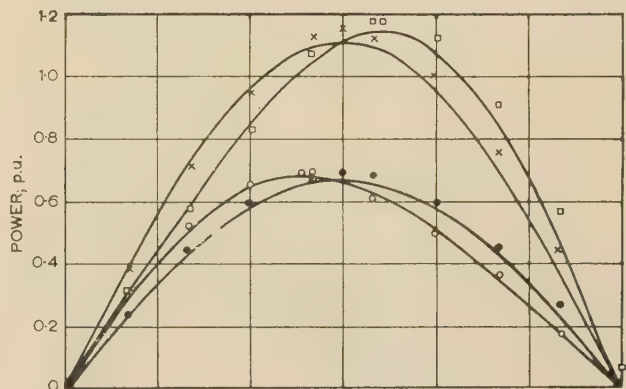
$$\text{Saliency factor} = 0.$$

(d) As in (c), but with transient saliency considered, i.e. Saliency factor = -0.429 .

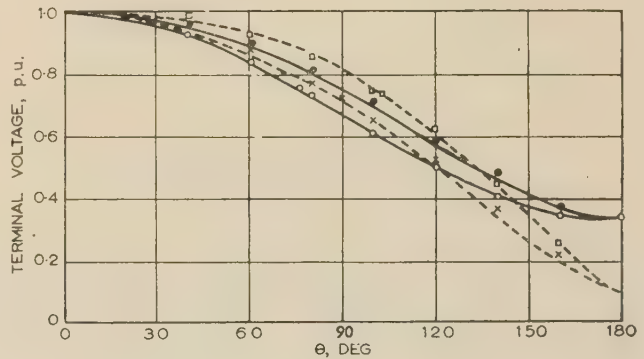
In cases (a) and (b) the e.m.f. was maintained constant at 1 p.u. In cases (c) and (d), the voltage behind the transient



(a)



(b)



(c)

Fig. 9.—Results by simulator and by calculation of a synchronous-machine problem.

- (a) Current/angle characteristics.
 (b) Power/angle characteristics.
 (c) Terminal-voltage/angle characteristics

The curves show the calculated values.
 The points show the simulator values as follows:

- Steady state, saliency neglected.
 ○ Steady state, saliency considered.
 × Transient state, saliency neglected.
 □ Transient state, saliency considered.

reactance was maintained constant at 1 p.u.; this implies an initial no-load condition for these two cases.

Figs. 9(a), 9(b) and 9(c) show the measured and calculated results for current, power and terminal voltage, respectively, against load angle. The calculated results were derived from the equations in Section 11. Comparison of calculated and experimental results is good except for the high values of measured power at load angles greater than 90° ; this is because the current is high at these values of angle, with consequent rapid increase

in the copper losses of the inductances representing the generator and line reactances.

(7) SIGNIFICANCE OF CONSIDERING SALIENCY

Reference to Figs. 9(a), 9(b) and 9(c) shows that the differences between the magnitudes of current, power and terminal voltage when saliency is considered and their magnitudes when saliency is neglected are of the order of 15, 20 and 15%, respectively. The figures of power and current are those for practical values of load angle, i.e. angles less than 30° . Also, the significance of the power divergence may be made clear by applying the equal-area criterion for the two cases; this is illustrated in Fig. 10,

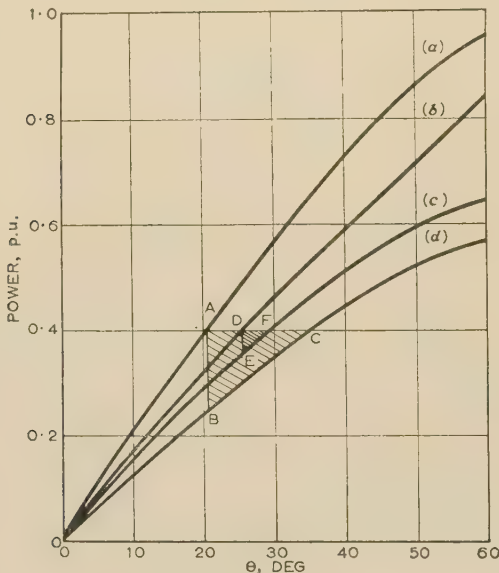


Fig. 10.—Rotor accelerating power.

- (a) Normal condition, saliency neglected.
- (b) Normal condition, saliency considered.
- (c) Faulted condition, saliency considered.
- (d) Faulted condition, saliency neglected.

which shows the areas proportional to accelerating power for an initial input power of 0.4 p.u. in the problem already considered. It may be seen from Figs. 9(b) and 10 that the curve of power transfer when saliency is neglected lies to the left of the corresponding curve when saliency is considered, the system being in the normal steady-state. When the system is faulted these curves are transposed, the saliency-considered curve now being to the left; this is due to the shift in phase of the second-harmonic component in the expression for power transfer. Thus, from Fig. 10, when saliency is neglected the accelerating power is proportional to the area ABC; when saliency is considered, the area is much reduced and is represented by the triangle DEF.

For the terminal voltage, the change in the two cases may well be less than 15% on the working portion. The significance of neglecting saliency may be better appreciated by noting the rate of change of voltage over the working range of load angle. In this case the slope of the terminal-voltage curve is approximately twice that of the corresponding curve when saliency is considered; this point is important under transient conditions if automatic voltage regulators are involved.

(8) CONCLUSIONS

From the results given in the paper, it is clear that the new simulator yields a result in accordance with calculation. The

phase-sensitive rectifier, the modulator and the overall feedback loop are accurate within 2, 1 and 3%, respectively; the discrepancy between measurement and calculation, over the working range, is 3% when the effect of copper losses in the inductance units has been taken into consideration. Some difficulty was experienced with the measuring system when the measured readings were small, a condition associated with increasing error. Under the worst possible conditions the total errors of the simulator were within 5%. It should be noted that this error exists with the same sign whether or not the effect of saliency is taken into account, and that it includes the errors of the network analyser. Reasons advanced for considering saliency are thus unaffected by questions of accuracy.

The simulator which has been described represents the steady-state saliency (and transient saliency on a steady-state basis) of synchronous machines on network analysers by a simple and automatic method. It is thus possible to eliminate the tedious calculation necessary whenever saliency is represented by a combined-reactance method on an a.c. network analyser. The results of the problem given demonstrate a simple case, with not unrealistic values of reactance, for which the effect of saliency is significant.

The differences between the new scheme described and those of Van Ness and Bauer are worthy of note. The apparatus here described is simpler than that of Bauer, having less equipment, fewer adjustments and thus presumably fewer sources of error; it is also simpler than that of Van Ness since it has no differential amplifier. The two earlier schemes use a d.c. signal to represent the e.m.f., whilst the apparatus under discussion represents the e.m.f. directly as an a.c. signal. A further point of importance from considerations of circuit arrangement is that, in the cases of Bauer and Van Ness, the input to the main feedback loop of their apparatus is a small signal proportional to current, whereas in the new method the signal is voltage-derived, much larger and, in consequence, easier to handle.

There is no difficulty in extending the range of the apparatus; for example, the addition of a phase-modulating arrangement, either electronic or electro-mechanical, will allow for the solution of stability problems.

It is estimated that the limited amount and cost of the apparatus involved do not preclude the addition of the direct representation of saliency to the generator units of large-scale network analysers.

(9) ACKNOWLEDGMENTS

The authors wish to acknowledge the advice of their colleagues in the Department of Electrical Engineering, Manchester College of Science and Technology, particularly that of Mr. B. D. Nellist.

(10) REFERENCES

- (1) DAVIES, M. W. H., and SLEMON, G. R.: 'Transformer Analogue Network Analysers', *Proceedings I.E.E.*, Paper No. 1479 S, October, 1953 (100, Part II, p. 469).
- (2) VAN NESS, J. E.: 'Synchronous Machine Analogues for Use with the Network Analyser', *Transactions of the American I.E.E.*, 1954, 73, Part III-B, p. 1054.
- (3) BAUER, H.: Discussion of 'System Stability, Load and Frequency Control', C.I.G.R.E., Paris, 1954, 1, p. 542.
- (4) CONCORDIA, C.: 'Synchronous Machines, Theory and Performance' (John Wiley and Sons, 1951).
- (5) MORTLOCK, J. R., and DAVIES, M. W. H.: 'Power System Analysis' (Chapman and Hall, 1952).
- (6) ADKINS, M. A.: 'Transient Theory of Synchronous Generators connected to Power Systems', *Proceedings I.E.E.*, Paper No. 1051 S, August, 1951 (98, Part II, p. 510).

- (7) DOHERTY, R. E.: 'A Simplified Method of Analysing Short-Circuit Problems', *Transactions of the American I.E.E.*, 1923, **42**, p. 841.
- (8) WAGNER, C. F.: 'Electrical Transmission and Distribution Reference Book', Chapter 6 (Westinghouse Electric and Manufacturing Company).
- (9) CRARY, S. B.: 'Power System Stability' (John Wiley and Sons, 1947).
- (10) KRON, G.: 'Equivalent Circuits of Electrical Machinery' (John Wiley and Sons, 1951).
- (11) TUCKER, D. G.: 'Balanced Rectifier Modulators without Transformers', *Electronic Engineering*, 1950, **22**, p. 139.

(11) APPENDIX

(11.1) Synchronous Machine connected to a Busbar through a Reactance

The case is considered of a synchronous machine with constant excitation, connected through an external reactance to a busbar of voltage V . The current in the two axes, neglecting resistance, is

$$I_d = \frac{E - V \cos \psi}{X_e + X_d}$$

$$I_q = \frac{V \sin \psi}{X_e + X_q}$$

The total current is

$$I = (I_d^2 + I_q^2)^{1/2}$$

The phase angle of the current at the busbar may be obtained easily from the expression

$$\phi = \arctan \left(\frac{I_d}{I_q} \right) - \psi$$

The terminal voltage of the machine may then be obtained from the equation

$$V_t = [(V + I X_e \sin \phi)^2 + (I X_e \cos \phi)^2]^{1/2}$$

The power is

$$P = V_d I_d + V_q I_q \\ = \frac{V E \sin \psi}{X_e + X_d} + \frac{V^2}{2} \left[\frac{1}{(X_e + X_q)} - \frac{1}{(X_e + X_d)} \right] \sin 2\psi$$

The angle for maximum power transmitted may be found from the equation

$$(\cos \psi)_{P_{max}} = \frac{1}{4} \left(-\frac{E (X_q + X_e)}{V (X_d - X_q)} \pm \left\{ \left[\frac{E (X_q + X_e)}{V (X_d - X_q)} \right]^2 + 8 \right\}^{1/2} \right)$$

All the equations above may be used for the case of transient studies on a steady-state basis, by replacing E and X_d by E'_q and X'_d respectively.

THE APPROXIMATE SOLUTION OF ELECTRIC-FIELD PROBLEMS WITH THE AID OF CURVILINEAR NETS

By L. TASNY-TSCHIASSNY, Dr.Tech.

(The paper was first received 23rd May, 1955, and in revised form 14th May, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

A rough field plot ascertained graphically can be used as the basis of a computational method of considerable accuracy. The plot is replaced by a net of orthogonal trajectories without sudden breaks of direction, the individual meshes being bounded by arcs of circles or straight lines. With the nodes of this net as the junction points of conductances, the 2-dimensional continuum within the problem boundaries can be converted into a network of lumped conductances. This network is analysed by measurement or computation.

The paper deals with the graphical and computational layout of a net whose contours follow the assumed trajectories closely, and explains the graphical and computational determination of the lumped conductances. A large number of formulae and computing routines are derived, making possible the employment of semi-skilled workers or computing aids. The determination of the gradient and other special problems, such as computing the areas associated with Poisson's equation or dealing with sharp corners, are discussed and the relevant formulae are given.

The advantages of the method, compared with the use of regular or irregular nets with straight-line inter-mesh boundaries, lie in the considerable reduction in the number of meshes for the same accuracy, and in the possibility of fitting the problem boundaries very closely. Two practical examples are given: using 4×6 meshes, the field gradient at the surface of an edge rounded by a non-circular curve is ascertained accurately within 0.6%; using 2×2 meshes, the cut-off frequency of the H_{11} mode in a circular waveguide is determined with an error of less than 1.6%; this is an eigenvalue problem.

(1) INTRODUCTION

In 2-dimensional field problems a graphical cut-and-try method of field mapping to determine the equipotentials and lines of force is frequently used by electrical engineers. In this method a curvilinear net approximating lines of force and equipotentials is drafted and methodically redrawn until the resulting map is considered satisfactory. The accuracy of the method is not great.

An entirely different approach to the solution of field problems consists in the replacement of the 2-dimensional continuum by a regular network of conductances and the analysis of the latter by computational methods¹ or measurements.² Irregularities in the conductance pattern must be introduced at the problem boundaries. The difficulties involved in this are avoided *a priori* by the employment of irregular nets⁴⁻⁶ instead of regular ones, or are overcome by special artifices.¹⁻³

A recent paper⁷ allows a combination of the two essentially different approaches. A rough field plot is drafted and approximated by parts of circles and straight lines in such a way that the orthogonality of the net is maintained, and sudden changes in direction of the field lines or contours in the mutually orthogonal families of curves do not occur. The resulting net is taken as the basis for the replacement of the 2-dimensional continuum

by a network of conductances. This electrical network can then be analysed by computation or measurement.

Practical experience with this method has demonstrated its suitability for engineering applications. One particular feature is the possibility of approximating irregular problem boundaries very closely, if not exactly, and this allows a higher accuracy for the values of voltage gradients at the boundaries. In addition, a considerable reduction in the number of meshes is possible for the same accuracy as that obtained with nets consisting of straight lines, which reduces the number of unknowns in the network to be analysed. This is of particular value if computing aids of limited size are employed.

A disadvantage of the method is the extra labour required in laying out the net and in computing the conductances. If the resulting network is analysed by computation, the relaxation patterns involved are more complicated than those for regular nets. This advantage is not very marked, since the linear forms that must be evaluated in the initial computation of residuals can be readily dealt with on desk calculators. For regular nets, similar complications appear at the boundary meshes only.

As shown in Reference 7, all advantages mentioned can be realized, although general computing procedures still had to be developed to create a practical engineering tool based on the method. The present paper gives an account of computing routines and procedures which are suitable for engineering purposes.

The basis of the method discussed is the well-known mathematical concept of conformal transformation, with the aid of which the 2-dimensional continuum within the problem boundaries can be approximated by a network of lumped conductances. Each mesh in such a network has to be dealt with individually, and the transformation equation used in each case depends on the type of mesh considered. This is discussed in Section 2, while Section 3 gives the analysis of the particular type of curvilinear mesh involved here, i.e. that bounded by parts of circles and straight lines. It is shown that such a mesh is always a part of a field map derived from a source and a sink of equal intensities but different polarities, and simple formulae can be derived for dealing with it. Section 4 provides the graphical, and Section 5 the computational, layout of an orthogonal net the meshes of which are bounded by parts of circles or straight lines or both, and which approximates a given rough field plot. Section 6 contains simple graphical and computational methods by means of which the conductances for the meshes in question can be computed. In Section 7 some special points of importance in practical work are discussed, and Section 8 gives two electrical engineering examples.

Mathematically, the paper deals with problems governed by the differential equation

$$\frac{\partial}{\partial x} \left(\sigma \frac{\partial V}{\partial x} \right) + \frac{\partial}{\partial y} \left(\sigma \frac{\partial V}{\partial y} \right) + J = 0 \quad \dots \quad (1)$$

This equation describes the current flow in a sheet of unit

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Tasny-Tschiassny, who was in the Electrical Engineering Department, University of Sydney, New South Wales, died on the 3rd August, 1955. The MS. was revised and the proofs were read by Mr. H. K. Messerle, Senior Lecturer in Electrical Engineering, University of Sydney.

thickness and conductivity σ , loaded by a transverse current density of $-J$. J is often zero, but it can be a function of position or of the potential V and its derivatives. The same differential equation applies to 3-dimensional axially symmetrical arrangements, provided that x denotes the radius and that the products of the actual values of σ and J with x are substituted for σ and J .

(2) THE CURVILINEAR NETWORK

Let us subdivide the domain within the problem boundaries by an orthogonal curvilinear net in such a way that sudden changes of direction of mesh boundaries do not occur and that changes in curvature of the mesh boundaries occur only at the nodes of the net. In this case a simple common analytical function for the transformation of the whole net into a rectilinear image does not as a rule exist, but simple functions for the individual meshes can be established. Since the meshes considered here have sides of constant curvature, only one transformation relation arises which applies to all meshes individually, and this fact simplifies the approach.

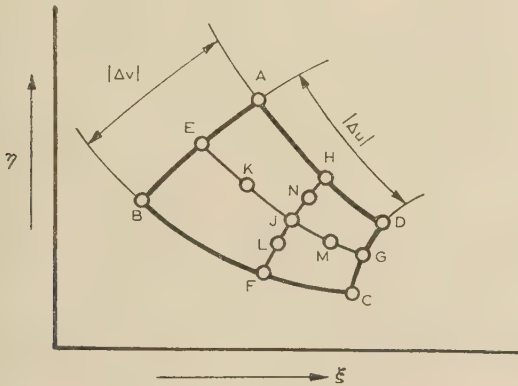


Fig. 1.—Curvilinear rectangle.

Considering any particular curvilinear rectangle $AZCD$ as shown in Fig. 1, we can define

$$w = u + jv = w(\zeta) = w(\xi + j\eta) \quad (2)$$

as the analytical function that converts the boundaries of this curvilinear rectangle in the (ξ, η) plane into a rectilinear rectangle in the (u, v) plane in such a way that the images of sides BA and CD are parallel to the v -axis and the images of sides AD and BC are parallel to the u -axis. As proved in Reference 7, the interior of the curvilinear rectangle $ABCD$ can be approximated by the arrangement of lumped conductances shown in Fig. 2, where $|\Delta u|$ and $|\Delta v|$ are the absolute values of the differences of the values of u and v for the relevant opposite sides. The points E, F, G, H, K, L, M, N and J result from a subdivision of the intervals Δu and Δv into four equal parts. The transverse loadings to be lumped at the points A, B, C and D correspond to the actual loadings of the rectangles $AEJH, BFJE, CGJF$, and $DHJG$ respectively.

The error involved in the values of the conductances is proportional to $[0(\Delta u^2) + 0(\Delta v^2)]$, and the multiplying factors of Δu^2 and Δv^2 are expressions of σ and V obtained by at least three differentiations with respect to u and v . The network shown in Fig. 2 can be replaced by the simple network shown in Fig. 3, if an error is permitted of $[0(\Delta u) + 0(\Delta v)]$ with multiplying factors involving at least two differentiations with respect to u and v . Owing to the nature of the multiplying factors, the errors become smaller, the better the boundaries of the curvilinear rectangle simulate equipotentials and their orthogonal trajectories.

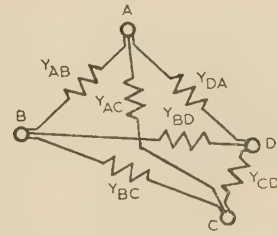


Fig. 2.—Replacement of the interior of a curvilinear rectangle by six conductances:

$$Y_{AB} = \frac{1}{2} \sigma_K \left| \frac{\Delta u}{\Delta v} \right| - Y_{AO}$$

$$Y_{BC} = \frac{1}{2} \sigma_L \left| \frac{\Delta v}{\Delta u} \right| - Y_{AO}$$

$$Y_{CD} = \frac{1}{2} \sigma_M \left| \frac{\Delta u}{\Delta v} \right| - Y_{AO}$$

$$Y_{DA} = \frac{1}{2} \sigma_N \left| \frac{\Delta v}{\Delta u} \right| - Y_{AO}$$

$$Y_{AO} = Y_{BD} = \frac{1}{2} \sigma_J \left(\left| \frac{\Delta u}{\Delta v} \right| + \left| \frac{\Delta v}{\Delta u} \right| \right)$$

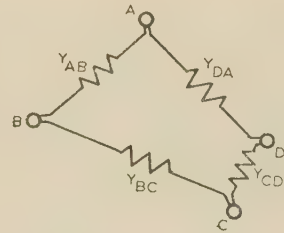


Fig. 3.—Replacement of the interior of a curvilinear rectangle by four conductances:

$$Y_{AB} = Y_{CD} = \frac{1}{2} \sigma_J \left| \frac{\Delta u}{\Delta v} \right|$$

$$Y_{BC} = Y_{DA} = \frac{1}{2} \sigma_J \left| \frac{\Delta v}{\Delta u} \right|$$

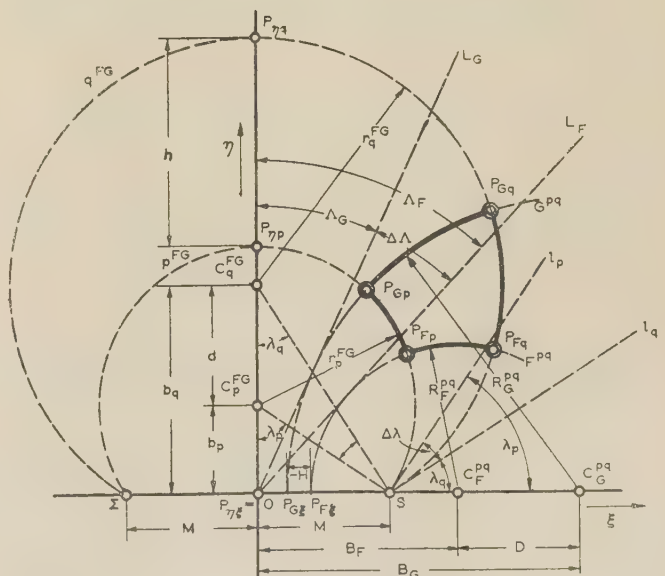


Fig. 4.—A curvilinear rectangle bounded by arcs of circles.

If the mesh is bounded by segments of constant curvature, the transformation function $w(\zeta)$ becomes fairly simple and will be discussed below. When considering a number of meshes we find the analysis still straightforward, although very tedious and repetitive. The notation used later can be deduced from Fig. 4, which shows a general curvilinear rectangle whose sides are parts of circles. Several characteristic designations are outlined here:

One family of contours is denoted by capital letters, e.g. F , G , etc.

The other family of contours, being orthogonal to the first, is denoted by small letters, e.g. p , q , etc.

Nodes are specified by subscript, e.g. P_{Fp} denotes the point of intersection of contours F and p .

Radii are denoted by one subscript and two superscripts, e.g. R_F^{pq} is the radius of curvature of the contour F between the contours p and q .

For simplification of indices the following symbols are used for the mesh defined by contours F , G , p and q :

$$Z = F \text{ or } G$$

$$n = p \text{ or } q$$

$$\text{e.g. } R_Z^{pq} = R_F^{pq} \text{ or } R_G^{pq}$$

$P_{Zn} = P_{Fp}$ or P_{Fq} or P_{Gp} or P_{Gq} , defining the four nodes of mesh $FGpq$.

(3) THE CURVILINEAR RECTANGLE BOUNDED BY SEGMENTS OF CONSTANT CURVATURE

The meshes considered here are curvilinear rectangles bounded by segments of constant curvature, i.e. the meshes are built up from arcs of circles and straight lines. The problem is to show how a mesh of this shape can be analysed.

As demonstrated in Fig. 4 and discussed below, the type of mesh to be investigated can be transformed into a rectilinear image by considering the mesh boundaries to correspond to parts of equipotential and flow lines in a field produced by a source and a sink of equal intensity but opposite polarity, being a certain distance apart. Such a field is made up of two families of circles, and the mathematics involved is well established.

A typical mesh as given in Fig. 4 is bounded by the contours F , G , p and q , which are sections of two equipotentials and two flow lines in the field produced by a source S and a sink Σ . Source and sink are located on the ξ -axis at a distance $\pm M$ from the origin O (see also Reference 7), and the field can be derived from the analytical function

$$w = u + jv = \log_{\epsilon} \frac{\zeta - M}{\zeta + M} = \log_{\epsilon} \frac{(\xi + j\eta) - M}{(\xi + j\eta) + M} \quad (3)$$

where u and v denote the potential and flow functions respectively.

Contours F and G are equipotentials, and the magnitude of the potential for each can be found by assuming the real part of the right-hand side of eqn. (3) to be constant. Thus, taking

$$u = \text{constant} = u_Z$$

the equipotential line Z is defined by a circle of radius R_Z^{pq} with its centre C_Z^{pq} lying on the ξ -axis at distance B_Z from the origin O . R_Z^{pq} and B_Z are given by

$$u_Z = \text{arc sinh} \left(\frac{M}{R_Z^{pq}} \right) \quad (4)$$

$$B_Z^2 = R_Z^{pq2} + M^2 \quad (5)$$

which represents a family of u_Z circles. Hence, if the radius of curvature of contours F and G is given, it is possible to find u_F and u_G .

The flow lines can be determined similarly by assuming

$$jv = \text{constant} = jv_n$$

This leads to a family of circles with all circles passing through S and Σ . The radius of the circles, r_n^{FG} , is given by

$$v_n = \text{arc sin} \left(\frac{M}{r_n^{FG}} \right) \quad (6)$$

and the centre lies on the η -axis at a distance b_n from the origin O , where

$$b_n^2 = r_n^{FG2} - M^2 \quad (7)$$

It follows, then, that the values of $|\Delta u|$ and $|\Delta v|$ for the curvilinear rectangle bounded by the contours F , G , p and q can be computed with the aid of eqns. (4) and (6), since each contour is part of a circle. The interior of the rectangle can be replaced by lumped conductances, as shown in Fig. 2 or 3. Some care is necessary, because arc sin is a multivalued function. The routine described in Section 6.2 shows how to avoid errors in this connection.

The analysis of mesh $FGpq$ can be generalized, and applies to any mesh of the type considered, and the appropriate ξ - and η -axes are always uniquely determinable. This can be shown by inspection of Fig. 4, whence it follows that the ξ - and η -axes are the common radical axes of the u_Z and v_n circles respectively.⁸

(4) THE GRAPHICAL LAYOUT OF A NET WITH SEGMENTS OF CONSTANT CURVATURE AS MESH CONTOURS

The problems arising in the practical layout of an orthogonal net with mesh boundaries composed of segments of constant curvature without sudden break of direction can in a general way be described with the aid of Fig. 5. Let A , E , and a , e be

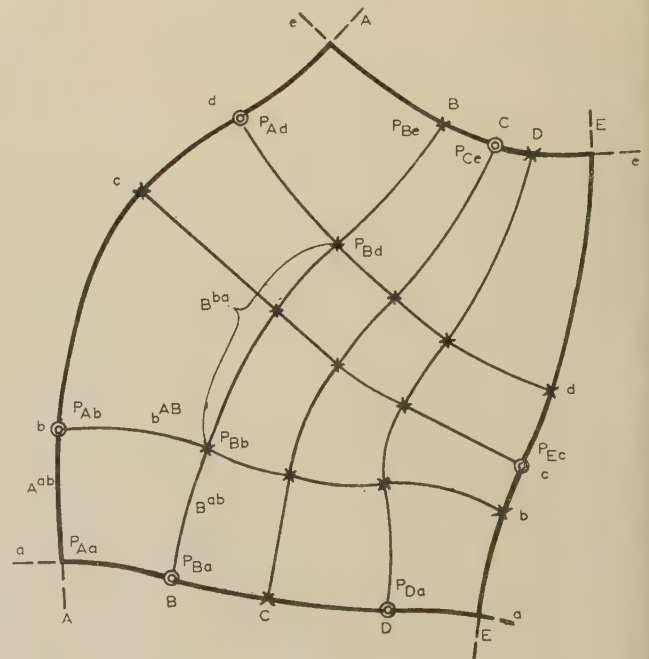


Fig. 5.—Laying out of a net.

the mutually perpendicular problem boundaries. Let the points P_{Ab} , P_{Ad} , P_{Ec} , P_{Ba} , P_{Da} , and P_{Ce} , marked by two concentric circles, be the points of break of curvature, if the boundaries are simulated by segments of constant curvature. Let a very

Part A:

(b) Select one boundary corner, e.g. the point P_{Aa} , as the starting corner.

(c) Assuming a suitable radius, simulate the part Bab of the contour B near the point P_{Ba} by a segment of constant curvature. Find the point of intersection P_{Bb} between this segment Bab and the curve of constant curvature bAB that is orthogonal to both Aab and Bab and passes through the point P_{ab} .

(d) Proceed similarly along the next part of the curve B , lay out the segments Bbd and dAB , and find the point P_{Bd} . Thereby the existence of subsidiary contours is ignored, e.g. of the contour c for which the points of break of curvature lie on the principal boundary E , and not on the principal boundary A .

(e) Continue in a similar way until the last subsidiary contour before the subsidiary boundary e is reached. In the example this is the contour d and was reached in step (d).

(f) From the penultimate point P_{Bd} draw the segment Bd^e perpendicular to both the segments d^{AB} and e^{AB} , and find the point of intersection P_{Be} on the subsidiary boundary e .

Part B:

To lay out the contour C a procedure similar to that in Part A is carried out. This procedure starts from that subsidiary boundary on which the next point of break of curvature lies, i.e. in the example it starts at the point P_{Ce} .

Part C:

Continuing as in Part B, lay out all principal contours and those subsidiary contours for which the break of curvature lies on the principal boundary A .

Part D:

Starting for each of the remaining subsidiary contours on the opposite principal boundary, i.e. in the example on boundary E , determine the run of the remaining subsidiary contours.

It can happen that a segment of an inner contour cannot satisfactorily be simulated by one straight line or one circle arc and that it is necessary to break its curvature at a convenient point. This does not affect the layout procedure. The only consequence is the addition of an extra inner contour of the other family through the internal point of break of curvature. For the layout of the additional contour we start from the internal point of break of curvature. Furthermore, it may be necessary to add internal contours to make the net finer without moving them ending at, or passing through, a point of break of curvature.

(5) THE COMPUTATIONAL LAYOUT OF A NET

For an accuracy greater than about 2% in the values of $|\Delta u|$ and $|\Delta v|$, computations must be used. Although it is possible to extend the computations to all meshes, as in the examples in Section 8, considerable improvement in accuracy can be achieved by limiting the computations to parts of the field, e.g. those with large potential gradients. Table 1 shows the part, referring to the mesh $FGpq$, of a table that has been found useful in the systematic attack on the layout of a net. The Table contains the most important data required for proceeding in accordance with the routine described in Section 5.5. A universal system of Cartesian co-ordinates (x, y) is chosen: x_{Zn} and y_{Zn} denote the co-ordinates of the point P_{Zn} , which could, for example, be P_{Fn} in Fig. 4.

$\Theta_{Zn}(-90^\circ < \Theta_{Zn} \leq 90^\circ)$ is the angle, with respect to the positive x -axis, of the normal (outward or inward, as the case may be) to the contour Z at the point P_{Zn} . Θ_{Zn} is the angle by which the positive x -axis must be turned anti-clockwise to

PART OF A SUITABLE LAYOUT TABLE

		p		q	
F		x_{Fp}	X_F^{pq}	x_{Fq}	
		y_{Fp}	Y_F^{pq}	y_{Fq}	
		θ_{Fp}	$\alpha_{Fp}^{pq} R_F^{pq} \alpha_{Fq}^{pq}$	θ_{Fq}	
		β_{Fp}^{FG}		β_{Fq}^{FG}	
		r_p^{FG}		r_q^{FG}	
		β_{Gp}^{FG}		β_{Gq}^{FG}	
G		x_{Gp}	X_G^{pq}	x_{Gq}	
		y_{Gp}	Y_G^{pq}	y_{Gq}	
		θ_{Gp}	$\alpha_{Gp}^{pq} R_G^{pq} \alpha_{Gq}^{pq}$	θ_{Gq}	

coincide with the direction of the normal. X_Z^{pq} , Y_Z^{pq} , and X_n^{FG} , Y_n^{FG} (the latter two not being recorded in Table 1) are the coordinates of the centres of curvature C_Z^{pq} and C_n^{FG} , and R_Z^{pq} and r_n^{FG} are the respective radii of curvature (see Fig. 4). α_{Zn}^{pq} (equal to either 0° or 180°) is the angle to be added to Θ_{Zn} to obtain the direction, with respect to the positive x -axis, of the outward normal to Z at the point P_{Zn} when approaching from between p and q . Similarly, β_{Zn}^{FG} (equal to 0° or 180°) is the angle to be added to Θ_{Zn} to obtain the direction, with respect to the positive y -axis, of the relevant outward normal to n when approaching from between F and G .

Formulae will be used that hold whether the principal boundaries of a mesh are the u - or the v -curves. Let Γ characterize the direction, with respect to the positive x -axis, of the positive ξ -axis, i.e. of the straight line leading from C_{PQ}^F to C_{PQ}^G (Fig. 4). Let Φ_{Zn} be the angle, with respect to the positive ξ -axis, of the relevant outward normal to F^{pq} at the point P_{Zn} , and let ϕ_{Zn} be the angle, with respect to the η -axis, of the relevant outward normal to n^{FG} at the point P_{Zn} . Φ_{Zn} and ϕ_{Zn} are either equal or differ by 180° .

(5.1) Formulae for all Mesh Boundaries being Parts of Circles

From simple trigonometrical considerations we obtain

$$\left. \begin{aligned} X_Z^{pq} &= x_{Zn} - R_Z^{pq} \cos (\Theta_{Zn} + \alpha_{Zn}^{pq}) \\ Y_Z^{pq} &= y_{Zn} - R_Z^{pq} \sin (\Theta_{Zn} + \alpha_{Zn}^{pq}) \end{aligned} \right\} \quad (8)$$

$$\left. \begin{aligned} X_n^{FG} &= x_{Zn} + r_n^{FG} \sin(\Theta_{Zn} + \beta_{Zn}^{FG}) \\ Y_n^{FG} &= y_{Zn}^* - r_n^{FG} \cos(\Theta_{Zn} + \beta_{Zn}^{FG}) \end{aligned} \right\} \quad (9)$$

Since the outward normal to a line of constant curvature is always directed to the same side of it, we obtain

$$(\alpha_{F_n}^{pq} \pm \beta_{F_n}^{FG}) \pm (\alpha_{G_n}^{pq} \pm \beta_{G_n}^{FG}) = (\alpha_{F_q}^{pq} \pm \beta_{F_q}^{FG}) \pm (\alpha_{G_q}^{pq} \pm \beta_{G_q}^{FG}) \quad (10)$$

Further formulae evident from the definitions of the quantities concerned are:

$$\Gamma = \arctan \frac{Y_G^{pq} - Y_F^{pq}}{X_G^{pq} - X_F^{pq}} \begin{cases} +0^\circ, & \text{if } X_G \geq X_F \\ +180^\circ, & \text{if } X_G < X_F \end{cases} \quad (11)$$

$$\Phi_{Z_n} = \Theta_{Z_n} + \alpha_{Z_n}^{pq} - \Gamma \quad . \quad . \quad . \quad (12)$$

$$\phi_{Z_n} = \Theta_{Z_n} + \beta_{Z_n}^{FG} - \Gamma \quad . \quad . \quad . \quad (13)$$

where $-90^\circ < \arctan \theta \leq 90^\circ$. The always-positive distance D between C_F^{pq} and C_G^{pq} (Fig. 4) is given by

$$D = \frac{X_G^{pq} - X_F^{pq}}{\cos \Gamma} = \frac{Y_G^{pq} - Y_F^{pq}}{\sin \Gamma} \quad (14)$$

From Fig. 4 we obtain

$$B_Z = -R_Z^{pq} \cos \Phi_{Zn} - r_n^{FG} \sin \phi_{Zn} \quad (15)$$

$$b_n = R_Z^{pq} \sin \Phi_{Zn} - r_n^{FG} \cos \phi_{Zn} \quad (16)$$

and

$$D = B_G - B_F \quad (17)$$

$$d = b_q - b_p \quad (18)$$

From eqns. (5) and (17) we then obtain

$$\left. \begin{matrix} B_F \\ B_G \end{matrix} \right\} = \frac{R_G^{pq2} - R_F^{pq2}}{2D} \mp \frac{D}{2} \quad (19)$$

and similarly, from eqns. (7) and (18),

$$\left. \begin{matrix} b_p \\ b_q \end{matrix} \right\} = \frac{r_q^{FG2} - r_p^{FG2}}{2d} \mp \frac{d}{2} \quad (20)$$

A closer investigation of Fig. 4 redrawn eight times for one point P_{Zn} according to the possible combinations $B_Z \geq 0$, $b_n \geq 0$ and $\alpha_{Zn}^{pq} = \beta_{Zn}^{FG}$ or $\alpha_{Zn}^{pq} = \beta_{Zn}^{FG} + 180^\circ$ leads to

$$\left. \begin{array}{l} \text{for } \alpha_{Zn}^{pq} = \beta_{Zn}^{FG} \\ \quad \left. \begin{array}{l} \text{If } b_n > 0 : 180^\circ + \Phi_{Zn} = 180^\circ + \phi_{Zn} \\ \text{If } b_n < 0 : \Phi_{Zn} = \phi_{Zn} \end{array} \right\} \begin{array}{l} \arctan \left(\frac{B_Z}{b_n} \right) \\ - \arctan \left(\frac{R_Z^{pq}}{r_n^{FG}} \right) \end{array} \\ \\ \text{for } \alpha_{Zn}^{pq} = 180^\circ + \beta_{Zn}^{FG} \\ \quad \left. \begin{array}{l} \text{If } b_n > 0 : \Phi_{Zn} = 180^\circ + \phi_{Zn} \\ \text{If } b_n < 0 : 180^\circ + \Phi_{Zn} = \phi_{Zn} \end{array} \right\} \begin{array}{l} \arctan \left(\frac{B_Z}{b_n} \right) + \arctan \left(\frac{R_Z^{pq}}{r_n^{FG}} \right) \end{array} \end{array} \right\} \quad (21)$$

(5.2) Formulae for Some of the Mesh Boundaries being Straight Lines

Fig. 6 shows a principal circle Z^{pq} and a subsidiary circle n^{FG} . Let A_Z and a_n be the distances of the points of intersection $P_{Z\xi}$ and $P_{\eta n}$ of the circles Z^{pq} and n^{FG} with the axes ξ and η , respectively, from the point of intersection $P_{\eta\xi}$ of the axes η and ξ . We obtain

$$\left. \begin{matrix} x_{\eta n} = x_{\eta\xi} - a_n \sin \Gamma \\ y_{\eta n} = y_{\eta\xi} + a_n \cos \Gamma \end{matrix} \right\} \quad (22)$$

$$\left. \begin{matrix} x_{Z\xi} = x_{\eta\xi} + A_Z \cos \Gamma \\ y_{Z\xi} = y_{\eta\xi} + A_Z \sin \Gamma \end{matrix} \right\} \quad (23)$$

$$A_Z = B_Z + R_Z^{pq} \cos \Phi_{Z\xi} \quad (24)$$

$$a_n = b_n + r_n^{FG} \cos \phi_{\eta n} \quad (25)$$

$$B_Z = (X_Z^{pq} - x_{\eta n}) \cos \Gamma + (Y_Z^{pq} - y_{\eta n}) \sin \Gamma \quad (26)$$

$$b_n = -(X_n^{FG} - x_{Z\xi}) \sin \Gamma + (Y_n^{FG} - y_{Z\xi}) \cos \Gamma \quad (27)$$

The angles $\Phi_{Z\xi}$ and $\phi_{\eta n}$ are either 0° or 180° . If R_Z^{pq} is eli-

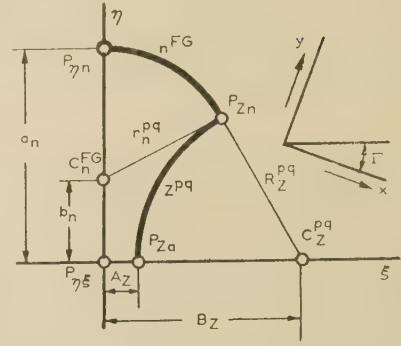


Fig. 6.—Mesh with straight boundaries.

minated from eqns. (5) and (24), and r_n^{FG} from eqns. (7) and (25), we obtain

$$B_Z = \frac{A_Z^2 + M^2}{2A_Z} \quad (28)$$

$$b_n = \frac{a_n^2 - M^2}{2a_n} \quad (29)$$

(5.3) Formulae for Two Mesh Boundaries being Concentric Circles and Two Straight Lines

The formulae given in Section 5.1 suffice to deal with this case.

(5.4) Formulae for all Mesh Boundaries being Straight Lines

$$\text{Here } \left. \begin{matrix} x_{Fq} - x_{Fp} = x_{Gq} - x_{Gp} \\ y_{Fq} - y_{Fp} = y_{Gq} - y_{Gp} \end{matrix} \right\} \quad (30)$$

$$\text{and } \theta_{Fp} = \theta_{Fq} = \theta_{Gp} = \theta_{Gq} \quad (31)$$

(5.5) The Routine of the Layout

The basis idea of the routine is to enable all computations to be carried out on the mental level of a computing machine without recourse to any drawing. Even the approximate description of such routine must be tedious, and the actual calculations are much simpler and quicker than they appear from their description. If a drawing is consulted, many steps of the routine become so obvious that they may not be realized as steps. Nevertheless, the existence of a foolproof routine is desirable, if semi-skilled personnel or machines are to be entrusted with carrying out the numerical computations.

The routine follows the procedure explained in Section 4, and the following are to be considered as given:

(a) The co-ordinates x and y , and the angles θ and α of all nodes to be followed up from the two principal boundaries and the radii of curvature of the segments between them.

(b) The co-ordinates x and y , and the angles θ and β of all nodes to be followed up from the two subsidiary boundaries and the radii of curvature of the segments between them.

(c) The radii of curvature of the segments of the principal inner mesh contours (with the exception of the last segments) and the indication (equality or inequality of the angles α pertaining to one point) where the curvature changes its sign.

(d) A schedule giving the order in which the individual meshes are to be examined.

The essential quantities to be ascertained when dealing with a mesh are the co-ordinates of the centre and the radius of curvature of the fourth, unknown, mesh contour and the co-ordinates and angles θ , α and β of the fourth, unknown, mesh node. Then an adjoining mesh can be examined.

Table 2

ROUTINE WHEN THE THREE GIVEN BOUNDARIES ARE PARTS OF CIRCLES

Step	Description
1	Compute X_F^{pq} , Y_F^{pq} , X_G^{pq} , Y_G^{pq} from eqn. (8) for $n = p$. Should the result be $X_G^{pq} = X_F^{pq}$ and $Y_G^{pq} = Y_F^{pq}$, the conditions of Section 5.3 are given.
2	Compute Γ from eqn. (11).
3	Compute Φ_{Fp} , Φ_{Gp} , Φ_{Fq} from eqn. (12). Note: If $\Phi_{Fq} = 0$ or $\Phi_{Fq} = 180^\circ$ (for $\Phi_{Fp} \neq 0$), q^{FG} is straight. Consult Table 3.
4	Compute ϕ_{Fp} , ϕ_{Gp} from eqn. (13).
5	Compute B_F , B_G from eqn. (15) for $n = p$.
6	Compute b_p from eqn. (16) for $Z = F$ or $Z = G$.
7	Compute $r_q^{FG} \sin \phi_{Fq}$ from eqn. (15).
8	Make $\phi_{Fq} = \Phi_{Fq}$ or $\phi_{Fq} = 180^\circ + \Phi_{Fq}$ to render r_q^{FG} positive. Compute r_q^{FG} . Compute β_{Fq}^{FG} from eqn. (13).
9	Compute b_q from eqn. (16) for $Z = F$.
10	Compute β_{Gq}^{FG} from eqn. (10) for $\alpha_{Gq}^{pq} = \alpha_{Gp}^{pq}$.
11	Compute Φ_{Gq} and ϕ_{Gq} from eqn. (21).
12	Compute θ_{Gq} from eqn. (12) or (13). If the condition $-90^\circ < \theta_{Gq} \leq 90^\circ$ is not complied with, add 180° to θ_{Gq} , α_{Gq}^{pq} , β_{Gq}^{FG} .
13	Compute x_{Gq} , y_{Gq} from eqn. (8).

Table 3

DEVIATIONS FROM TABLE 2 (EXCLUDING THOSE GIVEN IN THE TEXT), IF ONE MESH BOUNDARY, OR TWO MESH BOUNDARIES, EACH FROM A DIFFERENT FAMILY, ARE STRAIGHT

Straight contours	Step	Deviation
$F^{pq} = \eta$ ($\theta_{Fp} = \theta_{Fq}$)	7 9	Compute a_p from eqn. (25). Compute $x_{\eta\xi}$, $y_{\eta\xi}$ from eqn. (22) for $n = p$. Compute a_q from eqn. (22) for $n = q$. Find M^2 from eqn. (5) for $Z = G$. Find b_q from eqn. (29). Find $r_q^{FG} \cos \phi_{Fq}$ from eqn. (25). Omit.
$G^{pq} = \eta$ ($\theta_{Gp} = \theta_{Gq}$)	11, 12 13	Omit. Compute a_p and a_q from eqn. (25). Compute $x_{\eta\xi}$ and $y_{\eta\xi}$ from eqn. (22) for $n = p$. Compute x_{Gq} and y_{Gq} from eqn. (22) for $n = q$.
$p^{FG} = \eta$ ($\theta_{Fp} = \theta_{Gq}$)	5	Compute D from eqn. (14). Compute B_F , B_G from eqn. (19).
$q^{FG} = \xi$ (See Note to Step 3, Table 2)	7-9 10	Omit. Compute α_{Gq}^{pq} from eqn. (10) for $\beta_{Fq}^{FG} = \beta_{Gq}^{FG}$.
$\theta_{Fq} = \theta_{Gq}$	11, 12	Omit.
$F^{pq} = \eta$ $p^{FG} = \xi$ ($\theta_{Fp} = \theta_{Gp}$ $= \theta_{Fq}$)	5 7 9	Compute B_G from eqn. (26) for $n = p$. Compute a_q from eqn. (22). Compute M^2 from eqn. (5) for $Z = G$. Compute b_q from eqn. (29). Compute $r_q^{FG} \cos \phi_{Fq}$ from eqn. (25). Omit.
$G^{pq} = \eta$ $p^{FG} = \xi$ ($\theta_{Fp} = \theta_{Gp}$ $= \theta_{Gq}$)	5 11, 12 13	Compute B_F from eqn. (26) for $n = p$. Omit. Compute a_q from eqn. (25). Compute x_{Gq} and y_{Gq} from eqn. (22).
$F^{pq} = \eta$ $q^{FG} = \xi$ ($\theta_{Fp} = \theta_{Fq}$ $= \theta_{Gq}$)	7-12	Omit.
$G^{pq} = \eta$ $q^{FG} = \xi$ ($\theta_{Fp} = \theta_{Fq}$ $= \theta_{Gq}$)	7-12 13	Omit. Compute a_p from eqn. (25). Compute x_{Gq} and y_{Gq} from eqn. (22) for $n = p$.

Tables 2 and 3 give the routines for dealing with a mesh $FGpq$. The premisses are that all data are given referring to the points P_{Fp} , P_{Fq} , and P_{Gq} (with the exception of β_{Fq}^{pq}), and the radii of curvature of the segments F^{pq} , G^{pq} , and p^{FG} . The given data are shown in bold type in Table 1. The data referring to the segment q^{FG} and the node P_{Gq} are to be found.

Table 2 contains the standard procedure if the three given mesh boundaries are parts of circles. Suitable checks of the numerical calculations are: the values of b_p found in step (f) for $Z = F$ and $Z = G$ must be equal; and the values of D found from eqns. (14) and (17) must be equal, or the values of B_F and B_G computed from eqns. (14) and (19) must be equal to those found from eqns. (15).

Table 3 contains the deviations from Table 2 if some boundaries are straight, the simple cases given in Sections 5.3 and 5.4 being omitted. The application of the following rules is also omitted from Table 3 to save space:

- Do not use formulae computing or containing infinite values of radii or infinite values of co-ordinates of centres of curvature or infinite values of B_Z or b_n , i.e. infinite values pertaining to straight lines.
- For $F^{pq} = \eta$ or $G^{pq} = \eta$ make $\Gamma = \theta_{\eta p}$ or $\Gamma = 180^\circ \pm \theta_{\eta p}$, whichever is more convenient. Do not use eqn. (11).
- Make $\alpha_{Zp}^{pq} = \alpha_{Zq}^{pq}$ if $Z^{pq} = \eta$, and make $\beta_{Fq}^{FG} = \beta_{Gq}^{FG}$ if $n^{FG} = \xi$. It does not matter whether the values 0° or 180° are allotted to α_{Zp}^{pq} or β_{Fq}^{FG} .

(6) THE DETERMINATION OF $|\Delta u|$ AND $|\Delta v|$

After completion of the layout of the meshes, the values $|\Delta u|$ and $|\Delta v|$ required for the computation of the conductances according to Fig. 2 or 3 have to be determined. Methods will be used for determining $|\Delta u|$ and $|\Delta v|$ that are simpler than the twofold application of eqns. (4) and (6). These methods also cope successfully with the ambiguity of the arc-sin function.

(6.1) Graphical Determination of $|\Delta u|$ and $|\Delta v|$

From eqn. (6) it is evident that the angle λ_n (Fig. 4) measured in radians is equal to v_n . λ_n is also the angle between the tangent l_n to the circle n^{FG} at the point S and the positive ξ -axis. Consequently, $|\Delta v| = |v_q - v_p|$ equals the angle $|\Delta \lambda| = |\lambda_q - \lambda_p|$ between the directions l_q and l_p . Starting from eqns. (3), basic considerations show that this relation is correct whether the directions of the vectors $C_{Fp}^{FGP_{np}}$ and $C_{Fq}^{FGP_{nq}}$ coincide or differ by 180° , provided that the directions of the tangents l_p and l_q are taken along those circle arcs that lead to the points P_{np} and P_{nq} directly without passing through the point S.

It follows from eqn. (5) and Fig. 4 that the length of the tangent drawn from the origin $P_{\eta\xi}$ to any circle Z^{pq} equals M . Consequently, the angle Λ_Z between the η -axis and the tangent L_Z is given by

$$\tan \Lambda_Z = \frac{M}{R_Z} \quad (32)$$

or, using eqn. (4)

$$u_Z = \text{arc sinh} (\tan \Lambda_Z) = \text{amh}^{-1} \Lambda \quad (33)$$

where amh^{-1} denotes the functional relation called 'inverse Gudermannian'. Tables for amh^{-1} are available,⁹ and for comfortable work a protractor can be made from which $\text{amh}^{-1} \Lambda_Z$ is read off directly. $|\Delta u|$ is found as the difference of the relevant values u_Z . If the circles are on different sides of the η -axis, the absolute values $|u_F|$ and $|u_G|$ are to be added; if they are on the same side, their difference is to be taken.

Eqns. (32) and (33) break down if the two circles of one family are concentric. In this case we have

$$u_Z = \log_e R_Z^{pq} \quad (34)$$

so that $|\Delta u|$ is found as the difference of the values $|u_F|$ and $|u_G|$ ascertained from eqn. (34). $|\Delta v|$ equals the angle between the two straight lines into which the v -circles degenerate.

(6.2) Computational Determination of $|\Delta u|$ and $|\Delta v|$

A procedure is given that, without reference to a drawing, allows the computation of $|\Delta u|$ and $|\Delta v|$ from the data ascertained in the layout of the net. Let $|\Delta \Omega|$ be the value of $|\Delta u|$ or $|\Delta v|$ corresponding to the principal contours, and let $|\Delta \omega|$ be the value of $|\Delta v|$ or $|\Delta u|$ respectively that corresponds to the subsidiary contours.

(6.2.1) At Least One Principal and One Subsidiary Mesh Boundary is a Circle.

As shown in Section 9.1, $|\Delta \Omega|$ and $|\Delta \omega|$ can be computed by the following routines, as long as $|\Delta v|$ is less than π (180°), which is always complied with in practical work.

(6.2.1.1) All Mesh Contours are Curved.

Compute the quantities

$$K = \frac{\Phi_{Gq} - \Phi_{Gp}}{\Phi_{Fq} - \Phi_{Fp}} \quad (35)$$

$$k = \frac{\phi_{Gq} - \phi_{Fq}}{\phi_{Gp} - \phi_{Fp}} \quad (36)$$

where the signs of K or k indicate whether the two curves of a family are curved to the same or to opposite sides.*

Allot to R_G^{pq} and R_F^{pq} (or r_q^{FG} and r_p^{FG}) the same signs, if K (or k) is positive and different signs if it is negative. Select the sign of R_F^{pq} (or r_p^{FG}) in such a way that the smaller absolute value of H (or h) results from the equations†

$$H = (B_G + R_G^{pq}) - (B_F + R_F) \quad (37)$$

$$h = (b_q + r_q^{FG}) - (b_p + r_p^{FG}) \quad (38)$$

* For the purpose of eqns. (35) and (36) the angles $(\Phi_{Gq} - \Phi_{Gp})$ and $(\phi_{Gq} - \phi_{Fq})$ denote the change of the direction of the outer normal. Care is necessary if Φ or ϕ passes through 0° or 360° along the mesh boundary considered.

† H and h are the lengths shown in Fig. 4. If the signs of R_F^{pq} (or r_p^{FG}) are selected in such a way that H (or h) has the larger absolute value, the calculation, although correct, yields less accurate results. In Fig. 4 both r_p^{FG} and r_q^{FG} have the positive— in this case, the undesirable—sign.

Using the signs just ascertained for the quantities R_Z^{pq} , H , r_n^{FG} and h , compute

$$Q = 1 + \frac{H}{R_F^{pq}} - \frac{H}{R_G^{pq}} - \frac{1}{2} \frac{H}{R_F^{pq}} \frac{H}{R_G^{pq}} \quad (39)$$

$$q = 1 + \frac{h}{r_p^{FG}} - \frac{h}{r_q^{FG}} - \frac{1}{2} \frac{h}{r_p^{FG}} \frac{h}{r_q^{FG}} \quad (40)$$

Compute $|\Delta \Omega|$ and $|\Delta \omega|$ in the following way:

If Q or q is greater than 1:

$$\left. \begin{aligned} |\Delta Q| &= \text{arc cosh } Q \\ |\Delta \omega| &= \text{arc cosh } q \end{aligned} \right\}$$

If Q or q is between 0 and 1:

$$\left. \begin{aligned} |\Delta Q| &= \text{arc cos } Q \\ |\Delta \omega| &= \text{arc cos } q \end{aligned} \right\} \quad (41)$$

If Q or q is between -1 and 0 :

$$\left. \begin{aligned} |\Delta Q| &= \pi - \text{arc cos } |Q| \\ |\Delta \omega| &= \pi - \text{arc cos } |q| \end{aligned} \right\}$$

By arc cos is understood the angle in radians between 0 and $\pi/2$.

(6.2.1.2) One Mesh Contour is Straight, the Opposite One is Curved.

This may occur in one family only or in both families. A family in which the opposite mesh contours are curved is dealt with as described in Section 6.2.1.2. For the family in which one mesh contour is straight, compute Q' or q' , as the case may be, in the following way:

$$\text{If } B_Z \cos \Phi_{Z\xi} \geq 0; Q' = \mp \left| \frac{B_Z}{R_Z^{pq}} \right| \quad (42)$$

$$\text{and if } b_n \cos \phi_{\eta n} \geq 0; q' = \mp \left| \frac{b_n}{r_n^{FG}} \right| \quad (43)$$

In eqns. (42) and (43) Z and n refer to the curved contour of the family in which one contour is straight, and η and ξ are the axes that are members of the other family. The point $P_{Z\xi}$ (or $P_{\eta n}$) is that point of the two intersections with the axis ξ (or η) of the contour Z^{pq} (or n^{FG}), produced into a whole circle, i.e. in the case of a v -curve, reached from the point P_{Zn} without passing through S or Σ , and, in the case of a u -curve, nearer to the origin $P_{\eta\xi}$. In practice, $\cos \Phi_{Z\xi}$ (or $\cos \phi_{\eta n}$) will have the same sign as the two values $\cos \Phi_{Zn}$ (or $\cos \phi_{Zn}$).

The values of Q' and q' are used for the quantities Q and q in eqns. (41).

(6.2.2) Two Contours of One Family are Straight.

Let $\Delta \Lambda$ (or $\Delta \lambda$) be the angle between two intersecting opposite straight mesh contours that are principal (or subsidiary), and let r_q and r_p (or R_G and R_F) be the radii of the concentric circles that are the corresponding subsidiary (or principal) mesh contours. Then, as the case may be

$$|\Delta \Omega| = |\Delta \Lambda| \quad (44a)$$

$$|\Delta \omega| = \left| \log_e \frac{r_q^{FG}}{r_p^{FG}} \right| \quad (44b)$$

or

$$|\Delta \Omega| = \left| \log_e \frac{R_G^{pq}}{R_F^{pq}} \right| \quad (45a)$$

$$|\Delta \omega| = |\Delta \lambda| \quad (45b)$$

In this case

(7.1) Potential Gradients

At the point P_1 :

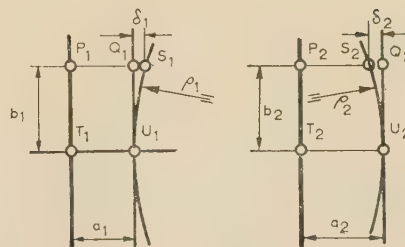


Fig. 7.—Diagram for correction in gradient computation.

At the point P_2 :

To find the value of ρ_1 or ρ_2 the following procedure suffices for the large radii of curvature that usually occur. The equal values of the potentials of the points T_1 and P_1 (or T_2 and P_2) are known, and so are the slightly differing values of the potentials at the points U_1 and Q_1 (or U_2 and Q_2). The distance δ_1 (or δ_2) defining the point S_1 (or S_2) that has the same potential as the point U_1 (or U_2) can be found by linear extrapolation (or interpolation). Then ρ_1 (or ρ_2) is given by

(7.2) Transverse Loading at Nodes

Fig. 8 shows three neighbouring principal mesh boundaries, F , G , H , and three neighbouring subsidiary mesh boundaries p , q , r . As mentioned in Section 2.1, the transverse loading by the current density, $-J$ [see eqn. (1)], is to be lumped at the

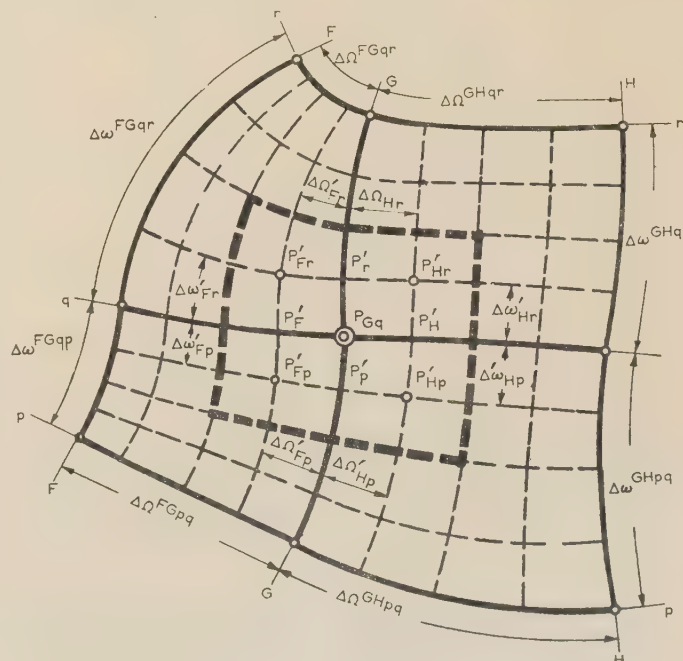


Fig. 8.—Subdivision of meshes to allow for transverse loading.

various nodes. With reference to the load to be lumped at the node P_{Gq} let

$$\left. \begin{aligned} \Delta\Omega'_{Fp} &= \frac{1}{4}\Delta\Omega'^{FGpq} \\ \Delta\omega'_{Fp} &= \frac{1}{4}\Delta\omega'^{FGpq} \end{aligned} \right\} \dots \dots \dots (53)$$

and let similar symbols be used for the other relevant combinations of the subscripts F , H , and p , r . Then four quarter interval points P'_{Fp} , P'_{Hp} , P'_{Fr} , and P'_{Hr} are determined. Let S'_{Fp} , etc., denote the area allotted to the point P'_{Fp} , etc. (see Section 2.1). We obtain for the current loading, $-I'_{Gq}$, to be lumped at the node P_{Gq} the value

$$I_{Gq} = S'_{Fp}J'_{Fp} + S'_{Hp}J'_{Hp} + S'_{Fr}J'_{Fr} + S'_{Hr}J'_{Hr} \quad (54)$$

where J'_{Fp} , etc., are the values of J at the points P'_{Fp} , etc. If the variation of J with the position does not differ very much from a linear variation, eqn. (54) can be approximately replaced by the more convenient form

$$\begin{aligned} I_{Gq} = & (S'_{Fp} + S'_{Hp} + S'_{Fr} + S'_{Hr})J_{Gq} + \Delta J'_p(S'_{Fp} + S'_{Hp}) \\ & + \Delta J'_r(S'_{Fr} + S'_{Hr}) + \Delta J'_F(S'_{Fp} + S'_{Fr}) + \Delta J'_H(S'_{Hp} + S'_{Hr}) \end{aligned} \quad (55)$$

In eqn. (55) $\Delta J'_p$, etc., denote the (algebraic) difference of the values of J'_p at the approximately defined points P'_p , etc. (Fig. 8) less the value of J_{Gp} at the point P_{Gq} . In many cases the first term in eqn. (55) is the only one that is significant. The computation of the areas S'_{Fp} , etc., can be sufficiently well approximated by the formula

$$S'_{Fp} = \frac{4\Delta\Omega'_{Fp}\Delta\omega'_{Fp}}{|dw/d\zeta|^2}, \text{ etc.} \quad (56)$$

where $\frac{1}{|dw/d\zeta|^2}$ is the Jacobian at the point P'_{Fp} , etc., of the $w - \zeta$ transformation (see, for instance, Reference 10) and is given by eqn. (49).

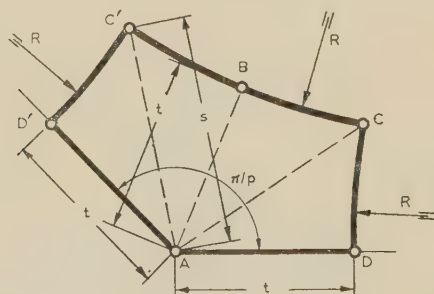


Fig. 9.—Sharp corner.

(7.3) Sharp Corners

Sharp corners formed by two straight lines meeting at an angle frequently arise as contours in engineering problems. The question of sharp corners was investigated in Reference 7. The most important result is given in Fig. 9 and by eqns. (57) and (58), namely

The neighbourhood of a sharp corner of aperture $(p\pi)$, bounded by two straight lines of arbitrary lengths t and three circles of radius R drawn symmetrically so that all adjacent mesh contours intersect each other at right angles (Fig. 9), can in very good approximation be replaced by two networks of conductances between the points $ABCD$ and $ABC'D'$ where the arrangements of the conductances correspond to those shown in Figs. 2 and 3 and their

values to those shown in Figs. 2 and 3 for $\Delta u = \Delta v = 1$. To meet the requirement of perpendicular intersection the conditions

$$R = \frac{t \sin(p\pi/4)}{1/\sqrt{2} - \sin(p\pi/4)} \quad (57)$$

$$S = \frac{t \sin[(1-p)\pi/4]}{1/\sqrt{2} - \sin(p\pi/4)} \quad (58)$$

which follow from geometrical considerations are to be fulfilled. The mesh contours shown in Fig. 9 can easily be made to fit into a normal pattern of curvilinear rectangles.

(8) EXAMPLES

Two simple examples will demonstrate the force of the use of curvilinear nets. So that a proper judgment might be made, no graphical constructions were used, and all computations were carried out to five digital places. We can expect the computational error in the conductance values and potentials to be not greater than 1 part in 500.

(8.1) Voltage Gradient at a Rounded Edge

The problem of the maximum voltage gradient at the inner contour of the arrangement shown in Fig. 10 (left) has been

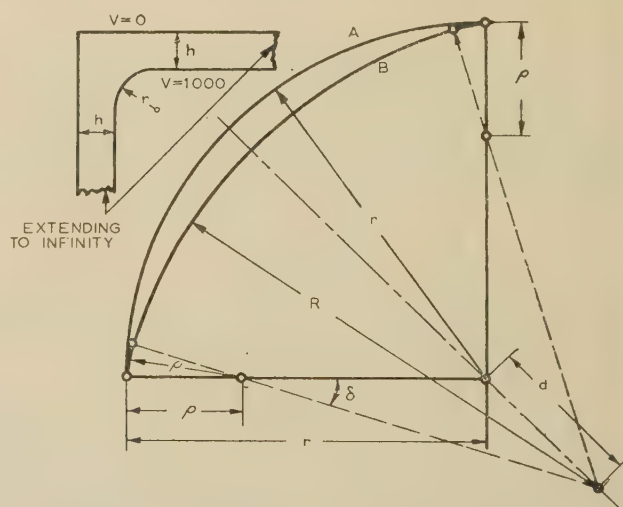


Fig. 10.—Rounded edge in the example in Section 8.1.

dealt with analytically by Cockcroft.¹¹ Using the method of conformal transformation he approximated the circular part of radius r of the inner contour by an analytically well-defined curve called a 'constant-stress curve', selected so that the voltage gradient along it is constant. Without any field plot, the constant-stress curve allows the computation of the maximum voltage gradient with the aid of algebra, the arc tan, and the logarithmic functions. The conditions for the ratio $r/h = 0.240$, for which computational results are available from Figs. 7 and 8 of Reference 11, will be investigated in two directions: First, the accuracy of the method described in the present paper will be checked by comparing the value of the maximum gradient ascertained by it with the value found by Cockcroft's formula; second, the error introduced by Cockcroft's approximation of the inner contour will be checked when the new method is applied to the circular and the constant-stress contours.

From Fig. 7 of Reference 11 we obtain for the gradient along the constant stress curve the value $\gamma = 1.92\Delta V/h$, where ΔV is the difference in potential between the contours. This value will be computed alternatively by a relaxation method. Using

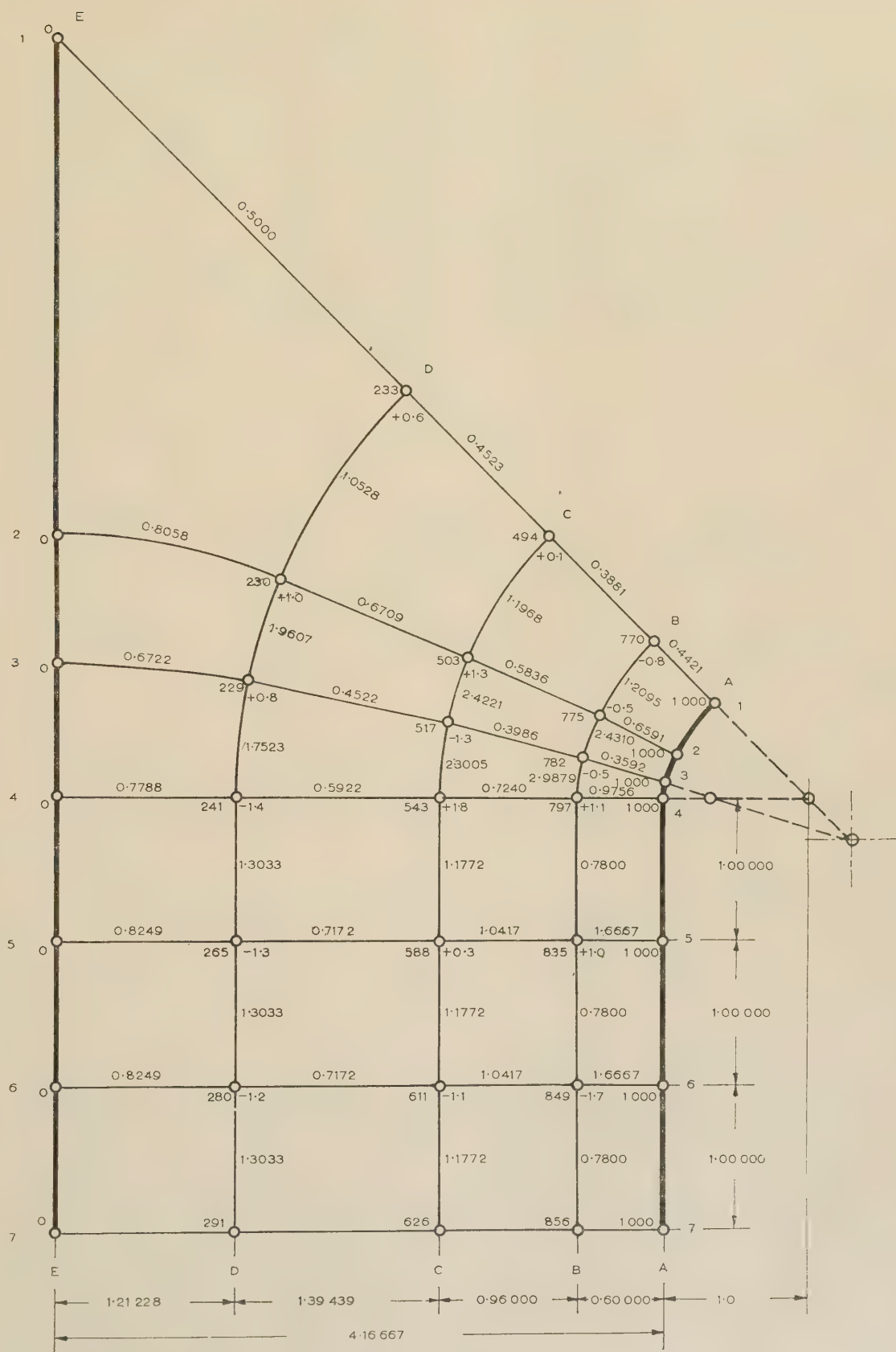


Fig. 11.—Example of net when using the constant-stress curve as inner boundary.

Fig. 8 of Reference 11 we find that the constant-stress curve can be simulated by three circles of radius R and ρ [see curve (b) in Fig. 10, right-hand side]. The original circle of radius r is represented by curve (a). Fig. 11 shows the layout of a suitable net. For reasons of symmetry, only one half of the corner need be dealt with, the line of symmetry '1' being a flow line. The point D2* of the sharp-cornered mesh DE12 is found by eqns. (57) and (58) from the condition that the radius of the arc D12 should coincide with the centre of the original inner arc (a) (Fig. 10). The contour D24 is the continuation of the arc D12. The contour C14 is an arc concentric with the arc D14 in an arbitrarily selected radial distance (1.39439 length units), whereas the arcs B13 and B34 are concentric with the arcs A13 and A34 given from Fig. 10 (right-hand side) in an arbitrary distance from them (0.600 length units). Selecting the point E7 as the origin of the general system of Cartesian co-ordinates (x, y) (Fig. 11), we obtain as premisses the figures given in bold type in Table 4.

conductances shown in Fig. 11 if the simpler network of Fig. 3 is taken as the basis throughout.

As boundary conditions the values 1000 for the boundary A, 0 for the boundary E, and the values corresponding to a uniform field distribution along the boundary 7 were assumed. The last assumption reduces the gradient slightly from that corresponding to the actual conditions, but the error should be negligible, judged by the comparison of one rounded corner associated with two infinitely long straight lines and two rounded corners at a finite distance.¹¹ The potentials ascertained by a relaxation procedure, and the residuals corresponding to them, are entered in Fig. 11.

For the gradient at the point A1 computed according to Section 7.1 we obtain either 1910/ h or 1908/ h according to whether the radius ρ_1 , appearing in eqn. (51a), of the curvature of the image in the w^{AB13} -plane of the line B13 is computed on the basis of the potential of the point B2 or the point B3. The correction factors for the curvature of the image of B13 are

Table 4
PART OF THE LAYOUT TABLE TO THE EXAMPLE IN SECTION 8.1

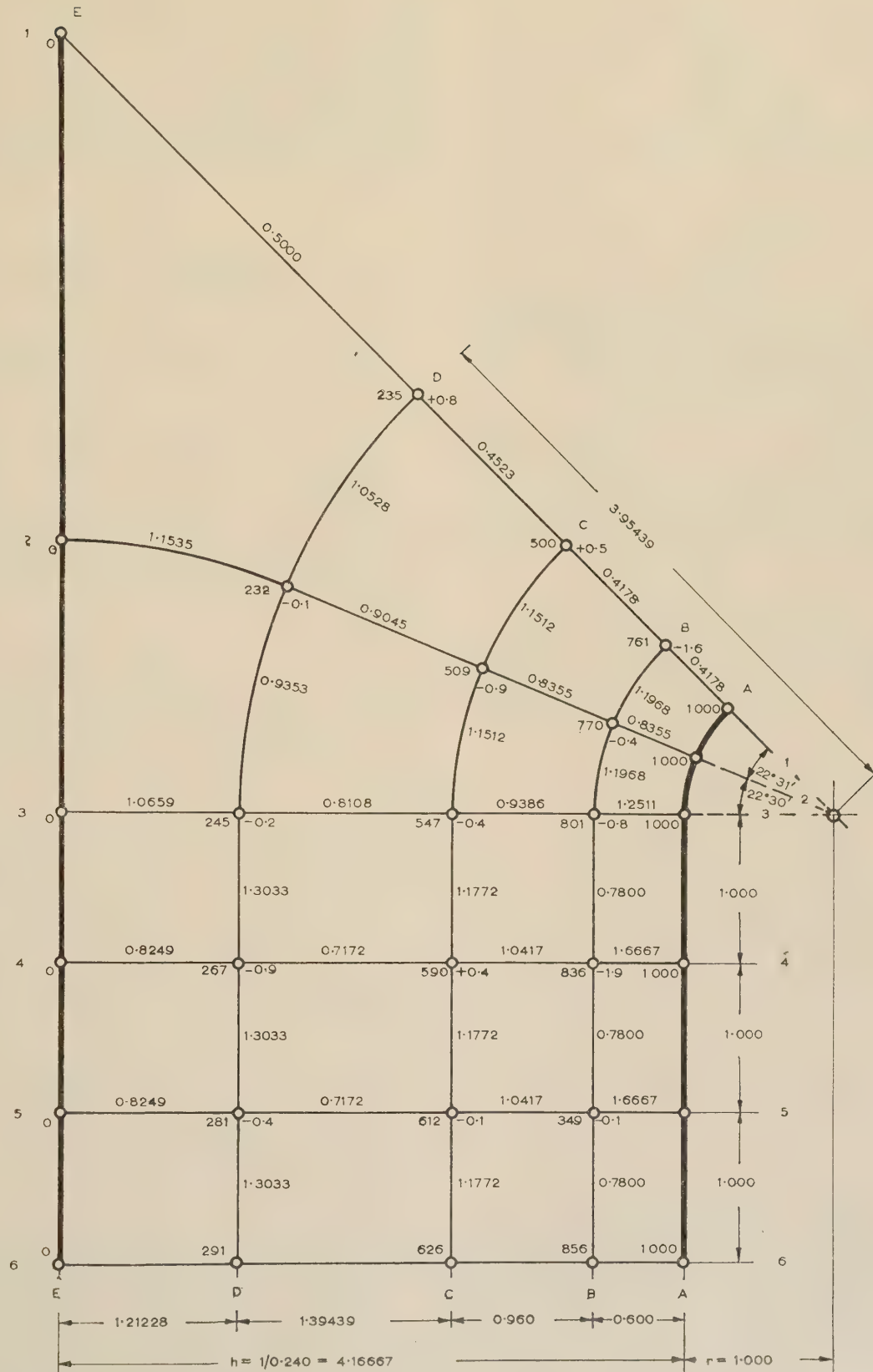
	1		2		3		4
A	-45°	5.47073 2.69594 180° 1.35000 180°	$-26^\circ 20' 5''$	5.47073 2.69594 180° 1.35000 180°	4.18100 3.094482 $-17^\circ 11' 1''$	4.48758 3.00000 180° 0.32091 180°	4.16667 3.00000 0°
	Straight		Straight		Straight		Straight
B	-45°	5.47073 2.69594 180° 1.95000 180°	$-26^\circ 20' 5''$	5.47073 2.69594 180° 1.95000 180°	3.60778 3.27209 $-17^\circ 11' 1''$	4.48758 3.00000 180° 0.92091 180°	3.56667 3.00000 0°
	Straight		0° 15.10189 0°		0° 10.09108 0°		Straight
C	-45°	5.16667 3.00000 180° 2.56000 180°	$-22^\circ 30'$	5.16667 3.00000 180° 2.56000 180°	2.65915 3.51571 $-11^\circ 37' 3''$	5.16667 3.00000 180° 2.56000 180°	2.60667 3.00000 0°
	Straight		Straight		Straight		Straight
D	-45°	5.16667 3.00000 180° 3.95439 180°	$-22^\circ 30'$	5.16667 3.00000 180° 3.95439 180°	$-11^\circ 37' 3''$	5.16667 3.00000 180° 3.95439 180°	1.21228 3.00000 0°
	Straight		3.95439		6.42015		Straight
E	0 Sharp angle	Straight	0 4.81430 0°	Straight	0 3.92822 0°	Straight	0 3.00000 0°

The selected net has the advantage that six meshes, namely AB12, AB23, AB34, CD23 and CD34, are formed by concentric circles and straight lines, and that for the remaining non-singular meshes only three different transformation functions are required, i.e. those governing the meshes BC13, BC34 and DE24. Owing to these advantages, the order of the computations required to fill in the remainder of Table 4 differs slightly from the standard order. The radii and centres of curvature of the principal curves (marked by capital letters) can be computed from simple geometric relations. Then, starting from A3, we compute the radii and angles connected with the line 3, and starting from E2 we compute those connected with the line 2. The data referring to the meshes inside the rectangle AE47 are easily computed. From the data ascertained we get the lumped

0.9907 and 0.9897 respectively. The maximum voltage gradient computed with the aid of the comparatively coarse curvilinear net is about 0.5–0.6% smaller than the theoretical value of 1920/ h . The error occurring when using the curvilinear net described is negligible for practical purposes.

To obtain an idea of the error introduced by replacing the circle (a) by the constant-stress curve (b), the approximate procedure was repeated with the net shown in Fig. 12. The contours B, C, and D were chosen as circles concentric with the boundary A, and the equivalent of the contour 3 in Fig. 11 was omitted. In Fig. 12 the resulting lumped conductances and the potentials and residuals found by a relaxation procedure were entered. For the gradient at the point A1 the value $\gamma = 2087.6k$ results, and the correction factor for curvature in the w^{AB12} -plane is 0.9853. Comparing the value 2087.6/ h gradient with the computed average value of 1909/ h for the gradient of the

* Instead of the notation P_{D2} , etc., the simpler notation D2, etc., will be used in the examples.



For the purpose of comparison, Table 6 also contains the accurate values of the potentials and the deviations from them. It is noteworthy that the deviations are greater than those occurring in Liebmann's method: his maximum deviations are ± 2.0 and -1.5% , resulting in a range of $\pm 1.75\%$, whereas the range in the present method is $\pm 4\%$. Nevertheless, the

cut-off wavelength is ascertained more accurately (0.82% compared with 2.2%). The smaller error in the cut-off wavelength found with the aid of a small number of curvilinear rectangles appears to be inherent in the method, and not accidental. If the error of -1.5% in the area is uniformly distributed, the wavelength increases by 0.75%, resulting in an error of about +1.6%. If the remaining terms in eqn. (55) are considered on the basis of an interpolation from the results in Table 6, a rough estimate shows that the error in the wavelength will be slightly reduced, although the errors in the values of the individual potentials may increase.

(9) REFERENCES

- (1) SOUTHWELL, R. V.: 'Relaxation Methods in Theoretical Physics' (Clarendon Press, Oxford, 1946).
- (2) LIEBMANN, G.: 'Precise Solution of Partial Differential Equations by Resistance Networks', *Nature*, 1949, **164**, p. 149. See also *British Journal of Applied Physics*, 1950, **1**, p. 92.
- (3) SHAW, F. S.: 'An Introduction to Relaxation Methods' (Dover Publications, Inc., 1953).
- (4) TASNY-TSCHIASSNY, L.: 'The Triangulation of a Two-Dimensional Continuum for the Purpose of the Approximate Solution of Second-Order Partial Differential Equations', *Journal of Applied Physics*, 1949, **20**, p. 419.
- (5) TASNY-TSCHIASSNY, L.: 'Asymmetrical Finite Difference Network for Tensor Conductivities', *Quarterly of Applied Mathematics*, 1955, **12**, p. 417.
- (6) MACNEAL, R. H.: 'An Asymmetrical Finite Difference Network', *ibid.*, 1953, **11**, p. 295.
- (7) TASNY-TSCHIASSNY, L.: 'Nets Composed of Parts of Circles for the Approximate Solution of Field Problems', *Australian Journal of Physics*, 1955, **8**, p. 8.
- (8) HALL, H. S., and STEVENS, F. H.: 'A School Geometry, Parts I-VI' (Macmillan, London, 1913), p. 336.
- (9) GREENHILL, A. G.: 'The Application of Elliptic Functions' (Macmillan, London, 1892), p. 16.
- (10) COURANT, R.: 'Differential and Integral Calculus' (Blackie, London, 1951), Vol. II, p. 535.
- (11) COCKCROFT, J. D.: 'The Effect of Curved Boundaries on the Distribution of Electrical Stress round Conductors', *Journal I.E.E.*, 1928, **66**, p. 385.
- (12) LIEBMANN, G.: 'The Solution of Waveguide and Cavity-Resonator Problems with the Resistance-Network Analogue', *Proceedings I.E.E.*, Monograph No. 38 R, May, 1952 (**99**, Part IV, p. 260).

(10) APPENDICES

(10.1) Derivation of Eqns. (35)-(41)

The angle $\Delta\lambda = \Delta v$, which is equal to the angle $C_p^{FG} S C_q^{FG}$ (Fig. 4) can be computed from the triangle $C_p^{FG} S C_q^{FG}$ by the cosine rule

$$\cos \Delta v = \frac{r_p^{FG} + r_q^{FG} - d^2}{2r_p^{FG}r_q^{FG}} \quad (61)$$

Let the radii r_n^{FG} be given algebraic signs, and let them be positive if the point of intersection P_{nn} of the segment considered of the circle n^{FG} with the η -axis is on the positive side of the centre C_n^{FG} of this circle. If in eqn. (61) the radii are substituted with their correct algebraic signs, the quantity Δv , so long as it is smaller than π , results correctly, provided that $0 < \Delta v < \pi/2$ or $\cos \Delta v > 0$, and $\pi/2 < \Delta v < \pi$ for $\cos v < 0$. If d in eqn. (61) is eliminated from eqn. (18) and then b_p and b_q are eliminated from eqn. (38) we obtain eqn. (40).

By using the expansion

$\cosh(\Delta u) = \cosh(u_G - u_F) = \cosh u_G \cosh u_F - \sinh u_G \sinh u_F$ eliminating $\cosh u_Z$ and $\sinh u_Z$ with the aid of eqns. (4) and (5), and by using eqns (17) and (37) with analogous sign conventions, we obtain first

$$\cosh \Delta u = \frac{P_F^{pq2} + R_G^{pq2} - D^2}{2R_F^{pq}R_G^{pq}} \quad (62)$$

and then eqn. (39). The relation between $\cosh \Delta u$ and $|\Delta u|$ is unique.

Since eqns. (61) and (62) have the same structure and can no doubt exist whether the function arc cos or arc cosh is to be used, no distinction need be made between the cases where the principal or subsidiary boundaries are the u -curves and the symbols $\Delta\Omega$ and $\Delta\omega$ can be introduced.

The rules in the text referring to eqns. (35) and (36) are the mathematical expression for the fact that the two values of R_n^{pq} or r_n^{FG} have equal or opposite signs. A simple algebraic check applied to eqns. (39) and (40) shows that a simultaneous change of the signs of both R_p^{pq} and R_q^{pq} or of r_p^{FG} and r_q^{FG} , and the corresponding change of the value of H or h , have no influence on the value of Q or q .

Eqns. (42) and (43) follow from the combination of eqns. (4) and (5) or (6) and (7), and from a check based on sketches of the possible combinations of the signs of B_Z or b_n and the angles concerned. The graphical analysis (Section 6.1), as well as a limiting procedure for $R_n^{pq} \rightarrow \infty$ or $r_n^{FG} \rightarrow \infty$ applied to eqns (39) and (40), lead to the same results.

(10.2) Approximate Solution of the Eigenvalue Problem of Eqn. (59) with the Aid of a Relaxation Method

Let V_r, V_s , etc., be the potentials of the nodes of the equivalent lumped network. By V_0 is understood the potential zero of the appropriate part of the boundary. Then a residual R_r can be expressed as the difference of two terms S_r and T_r , so that

$$R_r = S_r - T_r \quad (63)$$

where

$$S_r = \sum_{s=0}^{s=n} (V_r - V_s) G_{rs} \quad (64)$$

and

$$T_r = k_0^2 A_r V_r \quad (65)$$

G_{rs} is the lumped conductance connecting the nodes r and s , and A_r is the equivalent area pertaining to the node r , i.e. the area which, multiplied by V_r , corresponds to the loading of the node according to eqn. (55). Since even values of the potentials of the nodes that are approximate are not known in the first instance, we must start by neglecting all terms but the first in eqn. (55). A term $k_0^2 A_r$ can be considered as corresponding to a fictitious negative conductance connected between the nodes r and zero.

For a set of values V_r, V_s for which all residuals R_r are zero, k_0^2 can be found by multiplying all right-hand sides of eqns. (63) by the relevant potential V_r , adding the multiplied right-hand sides, and making them equal to zero. We obtain

$$k_0^2 = \frac{\sum S_r V_r}{\sum k_0^2 A_r V_r^2} \quad (66)$$

Let k_0^2 given by eqn. (66) be considered a function of arbitrary values of the potentials V_r . By differentiating k_0^2 with respect to all quantities V_r and making the partial derivatives equal to zero, we get eqn. (63) for all $R_r = 0$. Hence k_0^2 is stationary for slight deviations of the potentials V_r from their correct values. A recurrent method is now obvious: assume a suitable value of $(k_0^2)_{old}$; reduce the residuals as far as conveniently possible; and use eqn. (66) to find an improved value of $(k_0^2)_{new}$. For practical work it is more convenient to use eqn. (60), which can easily be derived from eqn. (66) when considering eqn. (63).

GOVERNING IN POWER SYSTEMS BY TIME-ERROR

By D. BROADBENT, Ph.D., B.Sc., M.Eng.Sc., Associate Member.

(The paper was first received 21st March, 1956, and in revised form 9th June, 1956. It was published as an INSTITUTION MONOGRAPH in September, 1956.)

SUMMARY

Governing in electric power systems by time-error provides a method of control with many advantages over speed governing, particularly if a high value of governor gain is used. The paper describes the operating results of a time-governed miniature system, the design of the governor for stable operation and the associated tie-line controllers.

LIST OF SYMBOLS

- K = Time-error governor gain, per-unit/rad.
 K_b = Stabilizing feedback gain.
 K_c = Tie-line angle controller gain, p.u./rad.
 K_t = Tie-line torque controller gain.
 M_l = Load torque.
 M_p = Prime-mover input torque.
 M_t = Tie-line torque transmitted.
 $p = \frac{d}{dt}$, time in radians.*
 T_a = Starting time of machine set, rad.
 T_b = Time-constant of stabilizing feedback, time-radians.
 T_d = Damping torque/speed change.
 T_{daz} = Damping torque/speed change between areas n and z .
 T_s = Electrical synchronizing torque coefficient, p.u./rad.
 T' = Steam header time-constant, time-radians.
 T'' = Servo-motor time-constant, time-radians.
 x_q = Alternator quadrature-axis synchronous reactance.
 x_t = Tie-line reactance.
 θ = Time-error angle.
 ϕ = Alternator torque angle.
 ζ = Damping factor.

Per-unit notation is used throughout.

(1) INTRODUCTION

(1.1) History of Governing

The speed governor is almost as old as the steam engine itself; Watt and Maxwell were both at different dates concerned in modifying the inherent regulation which an engine possesses, by some form of closed-loop system actuated by speed. The invention of Watt's pendulum governor was inevitable; the greater the power that is available the more exact must be its control if its application is to remain within the bounds set by materials, economy and knowledge.

As time passed it was this very exactness that caused difficulties and progress. A small error involves a high loop gain, and a high gain tends to give instability in any but a perfect system. Because of the imperfections of governed apparatus and valve gear, some governors later included dashpots which can now be recognized as lead-lag elements having a stabilizing effect.

* Time measured by the angular rotation in radians of a synchronous machine. For a 4-pole, 50 c/s machine, time in radians = $157 \times$ time in seconds.

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Broadbent is Senior Lecturer in Electrical Engineering, University of Melbourne, Australia.

Again, cases are known of hydro governors which were stable if belt-driven, but not if gear-driven. Maxwell in his early work dealt with both the throttle-speed and the friction-speed governor. The latter, in which friction is applied according to the speed error, can readily be seen to have a powerful stabilizing effect; a perfect speed governor, too, will damp oscillatory system performance. Such a governor having a typical per-unit (p.u.) regulation of 0.05 would be equivalent to a damping-torque coefficient of 20.0 (p.u. torque/p.u. speed change), in the system equation, which is a very large value indeed. Owing to time lags in the governor and elsewhere, nothing like this figure is realized in practice.

With the growth of electric power generation, the problem of tight control for economic operation was supplemented by that of the necessity to keep accurate electrical time. The first problem was dealt with very efficiently by the turbine engineers but without taking into account all the inherent regulating properties of an electrical synchronous system. The second problem has often been dealt with either by a clock-watching operator or by the engineers of the electricity supply undertaking installing an inching control operated from a clock on the governor-speed setting screw.

The problem, however, is not what supplementary controls need to be added to a speed governor to give it the correct transient and steady-state characteristics, but what is the best comprehensive control to ensure results consistent with good power-system operation in regard to frequency, tie-line and time control.

In electric power systems the speed governor has little to recommend it beyond the simplicity that has characterized it since its inception. Speed, but not its derivative or integral, can be momentarily judged by eye relatively easily for a rotating machine, and can be measured without difficulty by centrifugal force. On the other hand, a synchronous electric system has an inherent regulation associated with its torque angle, and since it is the electrical steady-state and transient errors that require modifying, the controlled variable should surely be the torque angle. It is partly for this reason that the governor should be controlled by time-error (small changes in time-error, measured as rotor angle, constituting changes in torque angle) and not the first time-derivative of rotor angle, which is speed. If stabilizing circuits³ are required, they may be designed from servo-mechanism theory calling for a second or third derivative term.

In January, 1952, the author constructed a time-error governor to operate without any stabilizing unit at a gain of 0.05 p.u. torque/radian time error. The error pick-off consisted of a synchro-differential which rotated according to the time-error angle between a voltage generated by a standard set and one generated by a pilot generator on the controlled set shaft. The rotation of the synchro-differential operated the throttle through a contactor servo-mechanism. The governor performance is described in Reference 1 and the principles of operation in a parallel synchronous system were described in an article² in February, 1953. Similar work was apparently being conducted by Électricité de France at this time, for in October, 1953, Cahen

and Chevallier published the results^{5,6} of tests using the very low gain of approximately 0.0003 p.u./rad. The time-error governors as used in the study to be described are completely electrical in nature. The error is represented by the out-of-balance voltage between synchro transformers on the shafts of the standard and controlled sets.

(1.2) Operating Principles of the Time-Error Governor

Fig. 1 is a vector diagram representing the rotors of a standard time set and two closely coupled sets, together with a vector S_{12} , representing the phase of the common stator voltage. The time-error angle θ is the angle by which the rotor of a particular set lags in relation to the standard, and ϕ is the torque angle by which the set rotor leads in relation to its generated voltage. From Fig. 1,

$$\theta_1 + \phi_1 = \theta_2 + \phi_2 \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Also, the power generated is given by

$$P_1 = T_{s1}\phi_1$$

and

$$P_2 = T_{s2}\phi_2$$

if ϕ is small and if T_s is considered a constant. If the turbine torque is made proportional to the time-error by the governor,

$$P_1 = K_1\theta_1$$

$$P_2 = K_2\theta_2$$

Substitution in eqn. (1) gives

$$\frac{P_1}{K_1} + \frac{P_1}{T_{s1}} = \frac{P_2}{K_2} + \frac{P_2}{T_{s2}}$$

For a load $(P_1 + P_2)$ is divided in the ratio

$$\frac{P_1}{P_2} = \frac{\frac{1}{K_2} + \frac{1}{T_{s2}}}{\frac{1}{K_1} + \frac{1}{T_{s1}}} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which is a constant and therefore indicates stable load distribution. (In a speed-governed system the load is divided inversely in the ratio of the governor regulations.) It is shown in Reference 2 that, by considering the transient vector diagram of an alternator, the relationship of eqn. (2) also holds in general for transient conditions. In all the systems constructed to date, including the one described here, T_s is much greater than K and, in consequence, load impacts are distributed according to the governor gain settings. In the interests of power-system stability studies, however, work is being continued by the author on the lines of making the value of K approach closely that of T_s . Stabilization and tie-line control, loads and alternator characteristics, all tend to complicate the analysis, but basically eqn. (2) still holds.

(1.3) Miniature Electric Power System

In investigating a new control, such as time-error governing, a miniature system has many advantages over other methods of determining system performance. It is sufficiently close physically to reveal the difficulties that might arise in the operation of a full-scale system, and the method of overcoming them in the miniature system can probably be repeated there. Again, because of its physical similarity, there is a possibility of making the so-called constants of the system vary in the same way as those of the parent system. Lastly, advocates of the speed governor can be convinced of the workability of time governing by a model. As a guide to the design of a miniature system, a series of

solutions to eqns. (3)–(9), but without the stabilizing terms, was obtained by means of a differential analyser.⁴ The main object was to arrive at an approximate value of the maximum gain at which a simple time governor could work and be stable, in a system having typical inertia, damping, header and servo-motor time lags. The miniature system, then, with properly designed stabilizing feedback, could be built round a considerably higher value of gain.

(2) DYNAMIC EQUATIONS AND OPERATION OF A SYNCHRONOUS SYSTEM

(2.1) Machine Set Time-Error-Governed

Throughout the paper small deviations from steady-state are considered, the quantity being prefixed by the symbol Δ . For this reason the coefficients of the equations may be taken as constants. For the isolated time-error-governed turbo-alternator set feeding a static load,

$$T_a p^2 \Delta\theta + T_d p \Delta\theta + \Delta M_p = \Delta M_l \quad . \quad . \quad . \quad (3)$$

where $\Delta\theta$ is measured against the continuously increasing rotor angle, and

$$\Delta M_p = \frac{K \Delta\theta}{(T'p + 1)(T''p + 1)} \quad . \quad . \quad . \quad (4)$$

if a simple time-error governor is used.

For optimum performance a stabilizing circuit is required, and, if a second derivative feedback is used with one predominant time-constant,

$$\Delta M_p = \frac{\Delta\theta}{(T'p + 1)(T''p + 1)} \left(K + \frac{K_b p^2}{T_b p + 1} \right) \quad . \quad (5)$$

where K_b depends on the gain of the stabilizing feedback network; T_b is the predominant time-constant of the stabilizing circuit in radians.

(2.2) Two Interconnected Synchronous Systems

If two turbo-alternator sets are connected through a synchronous tie each will affect the other to an extent depending upon the characteristics and strength of the tie. The equations for incremental load changes are:

$$T_{a1} p^2 \Delta\theta_1 + T_{d1} p \Delta\theta_1 + T_{d12} p (\Delta\theta_1 - \Delta\theta_2) + T_{s12} (\Delta\theta_1 - \Delta\theta_2) + \Delta M_{p1} = \Delta M_{l1} \quad . \quad (6)$$

$$T_{a2} p^2 \Delta\theta_2 + T_{d2} p \Delta\theta_2 + T_{d12} p (\Delta\theta_2 - \Delta\theta_1) + T_{s12} (\Delta\theta_2 - \Delta\theta_1) + \Delta M_{p2} = \Delta M_{l2} \quad . \quad (7)$$

where T_{d12} = Tie-line damping torque coefficient, i.e. the p.u. torque change through the tie-line per p.u. change in relative system speeds.

T_{s12} = Tie-line synchronizing torque coefficient, i.e. the p.u. torque change per radian of relative angular displacement of the two sets.

Subscripts 1 and 2 refer to the first and second machine set respectively.

The quantities in eqns. (6) and (7) are referred to one machine set as base.

Again, the terms ΔM_{p1} and ΔM_{p2} in eqns. (6) and (7) may be replaced by the respective governor transfer-function times $\Delta\theta$.

For time-error governing with stabilization

$$\Delta M_{p1} = \frac{\Delta\theta_1}{(T_1' p + 1)(T_1'' p + 1)} \left(K_1 + \frac{K_b p^2}{T_b p + 1} \right) \quad . \quad (8)$$

$$\text{and } \Delta M_{p2} = \frac{\Delta \theta_2}{(T_2'p + 1)(T_2''p + 1)} \left(K_2 + \frac{K_b p^2}{T_b p + 1} \right) \quad (9)$$

(2.2.1) Closely Coupled Governed Sets in Parallel.

If the synchronizing torque coefficient of an alternator T_s is very much greater than the governor gain K , the time-error angle θ is much greater than the torque angle ϕ and $\Delta \theta$ is also much greater than $\Delta \phi$, whence $\Delta M_{I1}/\Delta M_{I2} = K_1/K_2$ approximately, after a load impact $\Delta M_{I1} + \Delta M_{I2}$. This applies truly only when the stators are connected to the same busbar, in which case $\Delta \theta_1 = \Delta \theta_2$.

Eqns. (6), (7), (8) and (9) become

$$\left[T_{a1}p^2 + T_{d1}p + \frac{K_1 + \frac{K_b p^2}{T_b p + 1}}{(T_1'p + 1)(T_1''p + 1)} \right] \Delta \theta_1 = \Delta M_{I1} \quad (10)$$

$$\left[T_{a2}p^2 + T_{d2}p + \frac{K_2 + \frac{K_b p^2}{T_b p + 1}}{(T_2'p + 1)(T_2''p + 1)} \right] \Delta \theta_2 = \Delta M_{I2} \quad (11)$$

In eqns. (10) and (11) it is assumed that the rotors are effectively locked together at a fixed angle. The damping-torque coefficient T_d is associated then with the prime-mover and load damping only and is not affected by amortisseur windings on the alternator rotor.

(2.2.2) Machine Rotors in Free Oscillation.

It is also possible for the two rotors discussed in Section 2.2.1 to oscillate with each other almost independently of the governor action, more or less in free oscillation since the resulting frequency will be much higher than that of the governor oscillations.

On the assumption that the turbine throttle is fixed so far as these oscillations are concerned, the equation describing the system is:

$$(T_a'p^2 + T_{d12}'p + T_{s12})\Delta \theta_{12} = 0 \quad (12)$$

$$\text{where } T_a' = \frac{T_{a1}T_{a2}}{T_{a1} + T_{a2}}$$

and T_{d12}' is a composite damping coefficient.

The solution of eqn. (12) is:

$$\Delta \theta_{12} = Fe^{-(A/2)t} \cos \left[\sqrt{\left(B - \frac{A^2}{4} \right)} t + G \right]$$

$$\text{where } A = \frac{T_{d12}'}{T_a'} \text{ and } B = \frac{T_{s12}}{T_a'}$$

F and G depend on initial conditions.

The natural frequency of oscillation is $\omega_0 = \sqrt{B}$, and the actual frequency is $\sqrt{(B - A^2/4)}$, while the envelope curve of the damped oscillation has the form $Fe^{-(A/2)t}$.

Eqn. (12) may be rewritten in the standard form

$$p^2 + 2\zeta\omega_0 p + \omega_0^2 = 0$$

$$\text{where } \zeta, \text{ the damping factor, } = \frac{T_{d12}'}{2\sqrt{(T_a'T_{s12})}}$$

Standard non-dimensionalized curves may be used to read off directly the first and successive overshoots.

(2.2.3) Governed Set Operation when connected to an Infinite Busbar.

In Fig. 1, if S_{12} is a vector representing a machine stator and the phase of the infinite busbar voltage, the angle $(\theta + \phi)$ to the standard time reference is fixed because of the infinite nature of the system.

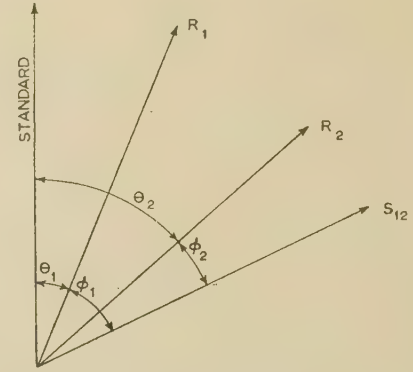


Fig. 1.—Rotating vectors representing standard time, machine rotors and alternator stator voltages.

The time-governed set can operate in steady state only in the condition in which $K\theta = T_s\phi$, and will be quite stable if the damping coefficients associated with oscillations of θ and ϕ are sufficiently large. $\Delta \theta = \Delta \phi$ in magnitude since S cannot oscillate with respect to the standard. Only one mode of oscillation will be apparent—one similar to the free electrical oscillation. Eqn. (6) becomes

$$\left[T_{a1}p^2 + (T_{d1} + T_{d1\infty})p + T_{s1\infty} + \frac{K_1 + \frac{K_b p^2}{T_b p + 1}}{(T_1'p + 1)(T_1''p + 1)} \right] \Delta \theta = \Delta M_{I1} \quad (13)$$

(2.2.4) Remotely Coupled Governed Sets in Parallel.

When one or more of the alternators is coupled to the remainder through a high impedance, e.g. a long tie-line, the synchronizing torque coefficient $T_s = [1/(x_q + x_d)]$ is reduced, and in consequence $\Delta \theta$ can approach $\Delta \phi$ in magnitude.

Fig. 2 shows the changes in angle of the rotors and point S of the system indicated when the load torque is increased by a small amount $\Delta M_{I1} + \Delta M_{I2}$. The diagram is drawn to represent

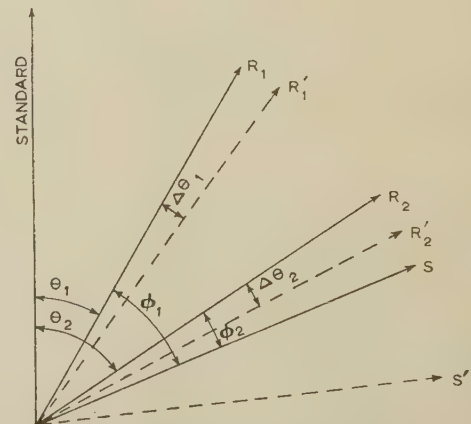
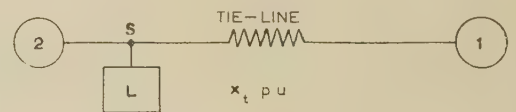


Fig. 2.—Rotor-angle changes in two time-governed systems connected by a tie-line when subjected to an increase of load.

approximately two sets working at the same load and governor-gain setting. While $T_{s1}\phi_1 = T_{s2}\phi_2$ and the ratio ϕ_1/ϕ_2 is therefore fixed by the electrical circuit, the ratio θ_1/θ_2 can be given any value by the throttle-setting control of the turbine, which varies the working torque independently of the working angle θ . This control is quite distinct from and independent of the governor-gain setting and corresponds to the speed-setting screw of the speed governor. It is essential if Fig. 2 is to be possible, because $K_1\theta_1 \neq K_2\theta_2$ if $K_1 = K_2$.

It can be seen that the voltage at point S after the load impact must move back through $\Delta\theta_1 + \Delta\phi_1$ from consideration of Set 1, and through $\Delta\theta_2 + \Delta\phi_2$ from consideration of Set 2.

$$\Delta\theta_1 + \Delta\phi_1 = \Delta\theta_2 + \Delta\phi_2 \quad \dots \quad (14)$$

$$\text{therefore} \quad \frac{\Delta M_{t1}}{\Delta M_{t2}} = \frac{\frac{1}{K_2} + \frac{1}{T_{s2}}}{\frac{1}{K_1} + \frac{1}{T_{s1}}} = \frac{K_1\Delta\theta_1}{K_2\Delta\theta_2}$$

in the steady state, or

$$\frac{\Delta\theta_1}{\Delta\theta_2} = \frac{1 + \frac{K_2}{T_{s2}}}{1 + \frac{K_1}{T_{s1}}} \quad \dots \quad (15)$$

The rotors can, however, still oscillate with each other as in sections 2.2.1 and 2.2.2. It is possible, then, that the two modes of oscillation, one due to governor and the other due to electrical torsional pendulum, are no longer independent.

(2.3) Tie-Line Controllers in a Time-Error-Governed System

Tie-line controllers and their application to conventional speed-governed systems have been described elsewhere.^{1,9} Tie-line control in a time-error-governed system raises no extra difficulties and again may be described as integral or proportional in operation.

(2.3.1) Integral Controller.

The modern tendency in power systems is to use the continuously acting integral type of tie-line power controller, sometimes called a floating control system, which ensures that the speed of the throttle-setting motor is proportional to the tie-line deviation. Since the valve position (and the turbine torque) depends on the position of the throttle motor shaft there results an integral type of controller having a describing equation of the form

$$p\Delta M_p'' = \frac{K_t\Delta M_t}{(T'p + 1)(T''p + 1)} \quad \dots \quad (16)$$

where $\Delta M_p''$ is the part of ΔM_p due to action by the tie-line controller and ΔM_t is the tie-line torque deviation.

(2.3.2) Proportional Controller.

A proportional type of controller may have a torque motor instead of the throttle-setting motor, and its shaft position (and hence M_p'') would depend on the tie-line deviation, thus

$$\Delta M_p'' = \frac{K_t\Delta M_t}{(T'p + 1)(T''p + 1)} \quad \dots \quad (17)$$

Such a controller is more stable generally than that described in eqn. (16) but has a steady-state position error associated with it. In other words, the tie-power will not be kept exactly at the required value but will deviate inversely as the controller gain.

(2.3.3) Angle-Sensitive Tie-Line Controller.

The angle-sensitive controller is peculiar to a time-error-governed system, and is discussed here because no description appears to have been published elsewhere. For simplicity it is assumed that the internal torque-angle deviation, $\Delta\phi$, of a machine is much smaller than the time-error-angle deviation, $\Delta\theta$, and can be neglected. Fig. 3 shows two machines, or areas, individually loaded and connected by a tie-line in which the torque transfer is to be kept constant.

It will be noticed that the reference vectors from which θ'_1 and θ'_2 are measured are displaced by angles ψ_1 and ψ_2 from the time standard; θ'_1 and θ'_2 are the angles as measured by the governor, while θ_1 and θ_2 are the true time-error angles. The ratio ψ_1/ψ_2 is not important, but $\psi_1 + \psi_2$ is the angle between tie-line ends for the approximations taken and for the special case shown in which $\theta'_1 = \theta'_2$. Variation of ψ_1 or ψ_2 acts as a throttle-setting control in the effect on θ'_1 and θ'_2 and the consequent response of the governors of turbines 1 and 2. The variation in tie-line torque can therefore be made to change ψ_1 or ψ_2 in a manner tending to reduce the variation. This is the tie-line-torque controller. In order that the controller can discriminate between an increase of load in one area and an equal decrease of load in the other, the controller is also made time-angle sensitive.

Eqs. (18) and (19) describe integral-type controllers at each end of the tie-line. Each controller sums the torque and angle signals and is phase-discriminating.

$$p\Delta M_{p1}'' = \frac{1}{(T'_1p + 1)(T''_1p + 1)}(K_{c1}\Delta\theta_1 + K_{t1}\Delta M_{t1}) \quad (18)$$

$$p\Delta M_{p2}'' = \frac{1}{(T'_2p + 1)(T''_2p + 1)}(K_{c2}\Delta\theta_2 + K_{t2}\Delta M_{t2}) \quad (19)$$

where $\Delta M_{t1} = -\Delta M_{t2}$.

For each value of tie-line-torque deviation there will be a value of time-angle deviation which will leave the controller unmoved.

Under such quiescent conditions,

$$p\Delta M_p'' = K_c\Delta\theta + K_t\Delta M_t = 0$$

$$\text{whence} \quad \frac{\Delta M_t}{\Delta\theta} = \frac{K_c}{K_t} \text{ p.u./rad, numerically.}$$

It can be seen that if the time-error governor gain of, e.g., area 2, K_2 , is made equal to K_{c2}/K_{t2} , the controller of area 2 will not operate for a load change in area 1, although $\Delta\theta_2$ will change momentarily, but the controller of area 1 will operate the throttle of machine 1 until the load has been absorbed, when $\Delta\theta_2$ reverts to zero.

Under these conditions the complete equations are:

$$(T_{a1}p^2 + T_{d1}p)\Delta\theta_1 = \Delta M_{p1} + T_{s12}(\Delta\theta_2 - \Delta\theta_1) + \Delta M_{t1} \quad (20)$$

where

$$\Delta M_{p1} = -\frac{K_1\Delta\theta_1}{(T'p + 1)(T''p + 1)} - \frac{K_b p^2 \Delta\theta_1}{(T'p + 1)(T''p + 1)(T_b p + 1)} - \frac{K_{c1}\Delta\theta_1}{p(T'p + 1)(T''p + 1)} + \frac{K_{t1}T_{s12}(\Delta\theta_2 - \Delta\theta_1)}{p(T'p + 1)(T''p + 1)} \quad (21)$$

and similarly for area 2.

Rearrangement of eqns. (20) and (21) gives

$$(T_{a1}p^2 + T_{d1}p + T_{s12})\Delta\theta_1 = \Delta M'_{p1} + \Delta M_{t1} + \left[T_{s12} + \frac{K_{t1}T_{s12}}{p(T'p + 1)(T''p + 1)} \right] \Delta\theta_2 \quad (22)$$

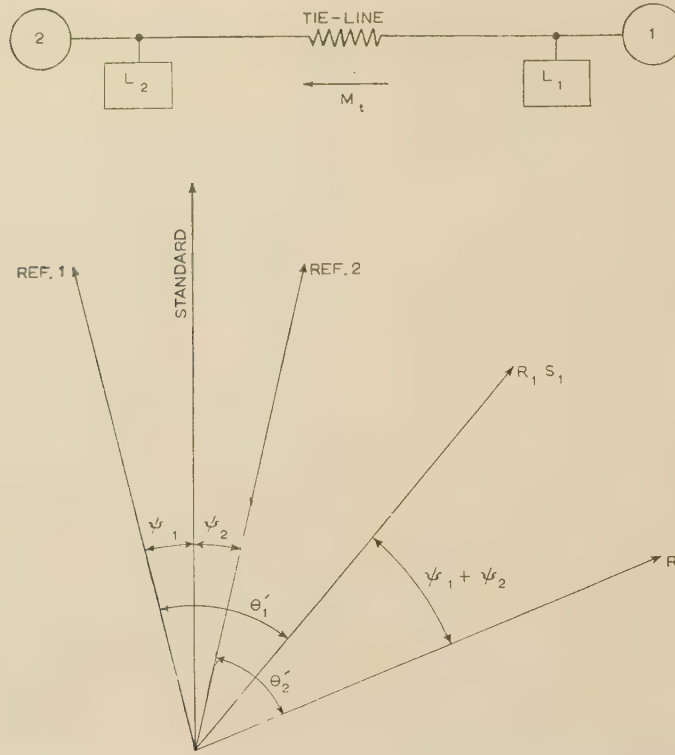


Fig. 3.—Two areas having their time governor references displaced from the time standard to allow for a large tie-line angle.

where

$$\Delta M'_{p1} = \left[-\frac{K_1}{(T'p+1)(T''p+1)} - \frac{K_b p^2}{(T'p+1)(T''p+1)(T_b p+1)} - \frac{K_{c1}}{p(T'p+1)(T''p+1)} - \frac{K_{t1} T_{s12}}{p(T'p+1)(T''p+1)} \right] \Delta \theta_1 \quad (23)$$

If K_{t1} is chosen so that for a 0.1 p.u. change in tie-power a prime-mover torque change of 1.0 p.u. follows in 30 sec (giving a somewhat higher value for K_{t1} than is used in conventional tie-line controllers),

$$p(\Delta M_{p1}) \text{ (from tie-line controller)} = \frac{1.0}{30 \times 157} = 0.1 K_{t1}$$

whence $K_{t1} = 0.002$.

Because of this low value of gain it is proposed to neglect the term $\frac{K_{t1} T_{s12}}{p(T'p+1)(T''p+1)}$ of eqn. (22) in comparison with T_{s12} , in spite of its integrating nature; as can be shown, it is unlikely to affect the stability of this particular system.

Eqn. (22) then becomes, for area 1,

$$\left(\frac{T_{a1}}{T_{s12}} p^2 + \frac{T_{d1}}{T_{s12}} p + 1 \right) \Delta \theta_1 = \frac{\Delta M'_{p1}}{T_{s12}} + \frac{\Delta M_{t1}}{T_{s12}} + \Delta \theta_2 \quad (24)$$

and for area 2,

$$\left(\frac{T_{a2}}{T_{s12}} p^2 + \frac{T_{d2}}{T_{s12}} p + 1 \right) \Delta \theta_2 = \frac{\Delta M'_{p2}}{T_{s12}} + \frac{\Delta M_{t2}}{T_{s12}} + \Delta \theta_1 \quad (25)$$

The block diagram of this system is shown in Fig. 4 while the feedback function $\Delta M'_{p1}/T_{s12}$ is shown in Fig. 5. Suggested values of the parameters and the gains of the controllers in decibels

(using, where possible, values used in practical power systems) are given below the diagram in Fig. 5; the two areas are taken as being identical. The forward elements are quadratic functions representing the basic oscillatory system where the natural frequency of oscillation $\omega_0 = \sqrt{(T_{s12}/T_d)}$ and the damping factor ζ is given by $2\zeta/\omega_0 = T_d/T_{s12}$. A reasonably strong tie is used with $T_{s12} = 0.2$, i.e. a 10% load transfer occurs when the electrical angle between tie-line ends is 1.0 electrical radian or 0.5 mechanical radian as used here.

The feedback elements $H_1 \dots H_4$ can influence the forward element G only when GH is greater than 0 dB. This is only the case for H_3 and H_4 and at frequencies so remote from the

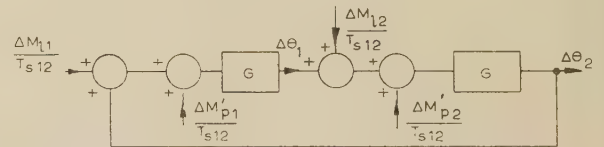


Fig. 4.—Block diagram representing two machines or areas coupled by a tie-line.

predominant mode of oscillation, ω_0 , that they will have little effect. If, on the other hand, any of the control elements (e.g. the governor gain K) had a gain approaching T_{s12} in value, the tie-line power oscillations would be influenced.

(2.3.4) Stabilization of Tie-Line Controllers.

Integral controllers which operate on the speed of the throttle-setting motor are likely to cause a response more oscillatory than their proportional counterparts. Accordingly, some authors⁹ suggest simple derivative stabilizing feedback for this type of controller. If, however, the servo-motor and steam-header time lags are considered (and they must be), simple derivative feed-

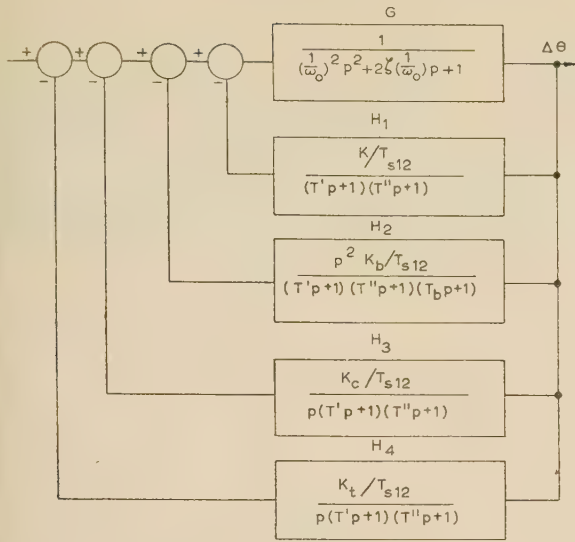


Fig. 5.—Modifications necessary to Fig. 4 if time governors and angle-sensitive tie-line controllers are used.

$$\begin{aligned} T_{s12} &= 0.2, & K/T_{s12} &= -9.1 \text{ dB.} \\ T_a &= 1050 \text{ rad.} & K_b/T_{s12} &= K_h T_d/T_{s12} = -15 \text{ dB.} \\ \omega_0 &= 2.16, & K_c/T_{s12} &= -43 \text{ dB.} \\ T_d &= 3.0, & K_t/T_{s12} &= -54 \text{ dB.} \\ \zeta &= 0.1, & T' &= T'' = 0.3 \text{ sec.} \\ K &= 0.07, & T_b &= 2.5 \text{ sec.} \\ K_t &= 0.002, \\ K_c &= KK_t = 0.0014. \end{aligned}$$

back has a limited stabilizing effect. An alternative scheme is put forward here and analysed.

If two areas connected by a tie-line are considered as a composite system, i.e. with composite inertia and damping, the angle between tie-line ends, θ_{12} , may be considered the variable and the tie-line power, $T_{s12}\theta_{12}$, the controlled quantity. If the governor and controller gains are of the same order as those given in Section 2.3.3, the tie-line power oscillations may be described by a second order system, thus

$$(T'_a p^2 + T_d p + T_{s12})\Delta\theta_{12} = \Delta M_l + \Delta M_p \quad (26)$$

$$\text{where } T'_a = \frac{T_{a1}T_{a2}}{T_{a1} + T_{a2}}$$

Eqn. (26) is represented by the forward element G_1 in Fig. 6.

$$G_1 = \frac{1}{\left(\frac{1}{\omega_0'}\right)^2 p^2 + 2\zeta' \frac{1}{\omega_0'} p + 1}$$

$$\text{where } \omega_0' = \sqrt{\frac{T_{s12}}{T'_a}} \quad \text{and} \quad 2\zeta' \frac{1}{\omega_0'} = \frac{T_d}{T_{s12}}$$

It is proposed to consider all the load and prime-mover torque deviations as zero, except for a part of the torque of one prime mover which is modified by the tie-line power as interpreted by the element H . If the parameters of Fig. 5 are used, $\omega_0' = 3.06 \text{ rad/sec}$ and $\zeta' = 0.15$; tie-line power swings are likely to be very oscillatory, with a first overshoot of over 1.6 p.u. and a frequency of 0.487 c/s.

A stabilizing feedback element H , as shown in Fig. 6, may be used, given by

$$H = \frac{0.28 p^2}{(T'p+1)(T''p+1)}$$

where $T' = T'' = 0.3 \text{ sec.}$

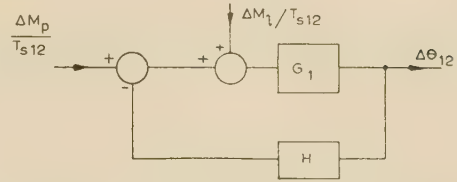


Fig. 6.—Block diagram representing two areas coupled by a tie-line considered as a composite system.

This element will be required to perform physically a triple differentiation on the tie-line power according to the equation

$$p(\Delta M_p \text{ tie-line stabilizer}) = \frac{K'_t p^3 \Delta\theta_{12}}{(T'p+1)(T''p+1)} \quad (27)$$

where $K'_t = 0.28/T_{s12} = 1.4$.

The resulting stabilized system has been solved for a step disturbance and the response is given in Fig. 7.

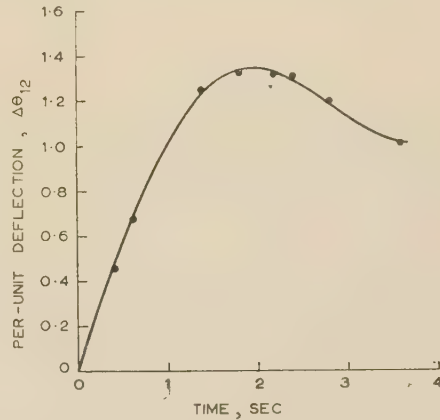


Fig. 7.—Tie-line power swings in the system shown in Fig. 6.

(3) RESULTS OF GOVERNED-SYSTEM OPERATION

Two machine sets were used to obtain the results given here. A step-function electrical load disturbance caused rotor oscillations; the change of angle between the rotors of the test machines and that of a third standard reference set were recorded on a high-speed pen recorder.

The oscillograms obtained are the practical interpretation of the equations in Section 2 which were used in the design of the tests. The steady-state load exchanges as predicted in Sections 1 and 2 are confirmed and the effect of varying the circuit parameters is shown. The acceleration time-constant T_a was kept constant at the low value of 1050 rad, which equals 6.7 sec, for both machines throughout the study. The values of the damping-torque coefficients T_d were of the same order as those associated with practical working conditions. The governor-gain constants K_1 and K_2 , although less than a machine synchronizing-torque coefficient, were considerably higher than anything contemplated to date and begin to approach the synchronizing-torque coefficients of longer tie-lines. The governor servo-motor time-constant T'' was kept constant throughout the study at 0.35 sec (55 rad) but the steam-header time-constant T' was made zero in some tests and approximately 0.3 sec (47 rad) in others. For the two sets in parallel two values of tie-line reactance were used, zero and 1.9 p.u. In all the tests of which the results are shown the governor incorporated a second derivative stabilizing feedback.

(3.1) Isolated Time-Error-Governed Set

Eqns. (3) and (5) describe the stabilized isolated governed set. Oscillograms (a), (b) and (c) in Fig. 8 show the effect of load

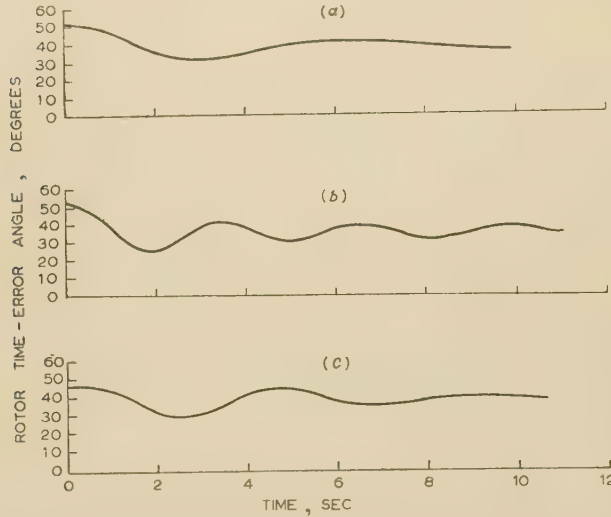


Fig. 8.—Record of $\Delta\theta_1$; Set 1 isolated; 0.85 p.u. load (1 kW base).

$$T_{a1} = 1050 \text{ rad}; T_{d1} = 3.5.$$

$$(a) K_1 = 0.08; T'_1 = 0 \text{ sec}; T''_1 = 0.35 \text{ sec}; \Delta M_1 = 0.014 \text{ p.u.}$$

$$(b) K_1 = 0.16; T'_1 = 0 \text{ sec}; T''_1 = 0.35 \text{ sec}; \Delta M_1 = 0.042 \text{ p.u.}$$

$$(c) K_1 = 0.08; T'_1 = 0.29 \text{ sec}; T''_1 = 0.35 \text{ sec}; \Delta M_1 = 0.014 \text{ p.u.}$$

disturbances ΔM_1 in the change of rotor or time-error angle; detailed information is given beneath the Figure. The effect of including the steam-header time lag is shown by comparing oscillogram (a) with oscillogram (c). The oscillations are more sustained in the case of machines having a steam-header time lag, but not seriously so for the values of K and T_d used. A doubling of the gain value [oscillogram (b)] has a greater effect in this respect and the frequency of oscillation is increased to approximately 0.33 c/s.

(3.2) Closely Coupled Time-Governed Sets in Parallel

Oscillograms (a) and (b) in Fig. 9 and (a)–(c) in Fig. 10 record the rotor time-error angle of one set after a total load disturbance of 4.2% was applied to the paralleled sets. Those in Fig. 9 are for no-load or synchronizing conditions and show the natural higher-frequency mutual rotor oscillations. Eqns. (6)–(9) describe the motions fully, but it is more convenient to separate the two predominant modes of oscillation and consider them independently. The difference in their respective frequencies of

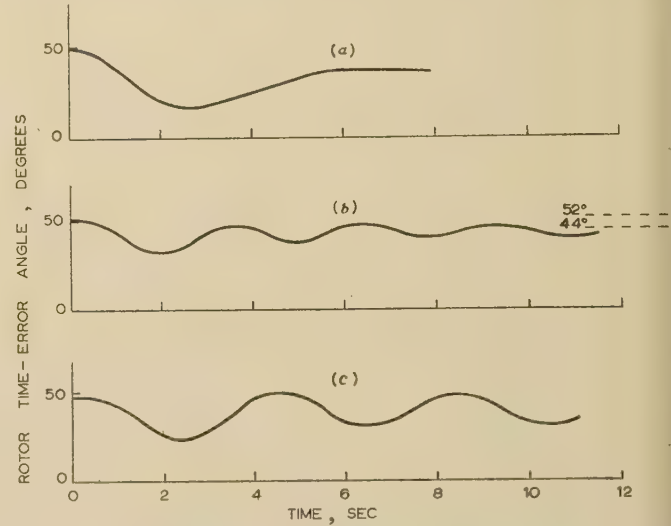


Fig. 10.—Record of $\Delta\theta_2$; Sets 1 and 2 in parallel; 0.85 p.u. and 0.93 p.u. load.

$$T_{a1} = T_{a2} = 1050 \text{ rad.}$$

$$T'_{d12} = 1.9.$$

$$T_{d1} = 3.5.$$

$$T'_a = 525.$$

$$T_{d2} = 3.1.$$

$$T_{s12} = 3.9.$$

$$(a) K_1 = 0.08; K_2 = 0.065 (0.07); T'_1 = T'_2 = 0.35 \text{ sec}; T'_1 = T'_2 = 0; \Delta M_1 = 0.042.$$

$$(b) K_1 = 0.16; K_2 = 0.13; T''_1 = T''_2 = 0.35 \text{ sec}; T'_1 = T'_2 = 0; \Delta M_1 = 0.042.$$

$$(c) K_1 = 0.08; K_2 = 0.07; T'_1 = T'_2 = 0.35 \text{ sec}; T'_1 = 0.29 \text{ sec}; T'_2 = 0.32 \text{ sec}; \Delta M_1 = 0.042 \text{ p.u.}$$

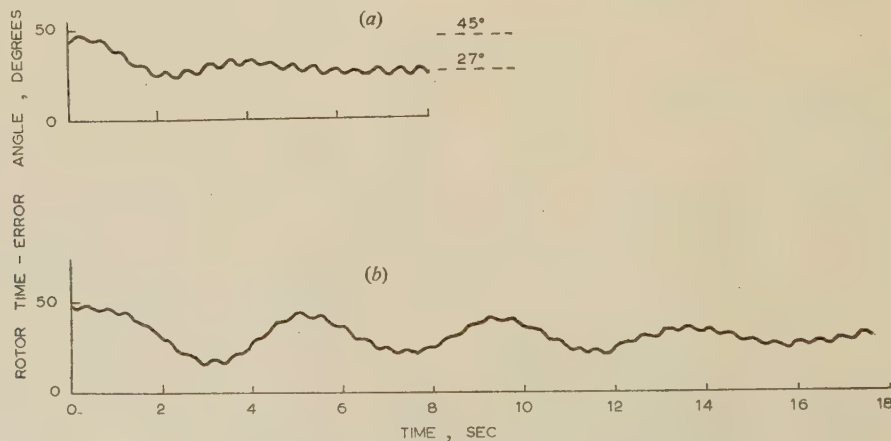


Fig. 9.—Record of $\Delta\theta_2$; Sets 1 and 2 in parallel; no-load.

$$T_{a1} = T_{a2} = 1050 \text{ rad.}$$

$$T'_a = 525 \text{ rad.}$$

$$T_{d1} = 2.7.$$

$$T'_{d12} = 1.6.$$

$$T_{d2} = 3.0.$$

$$T_{s12} = 3.5.$$

$$K_1 = 0.06, K_2 = 0.11.$$

$$x_t = 0.$$

$$T'_1 = T'_2 = 0.35 \text{ sec.}$$

$$\Delta M_1 = 0.042 \text{ p.u.}$$

Ripple frequency, 2.1 c/s.

$$(a) T'_1 = T'_2 = 0.$$

$$(b) T'_1 = 0.29 \text{ sec}; T'_2 = 0.32 \text{ sec.}$$

1 and approximately 0.25 c/s justifies this. The higher-frequency oscillation is according to eqn. (12) and is an inherent property of the electrical system.

The governor oscillations are according to eqns. (10) and (11) in which the electrical oscillations are assumed to have no effect. As the sets are similar the governor response of the other set is identical but its higher frequency deflections are the negative of those shown. The values of T_{d1} and T_{d2} are higher than those associated with a practical system during synchronizing, and a reduction of governor gain would be necessary in the interests of stability. Oscillograms (a)–(c) in Fig. 10 correspond to load

conditions and were continued long enough for both sets to attain the steady-state shaft swing angle, showing that the load distribution is proportional to the governor gains, in that $\Delta\theta_1 = \Delta\theta_2$ finally.

(3.3) Remotely Coupled Time-Governed Sets

The oscillograms in Figs. 11 and 12 show the rotor oscillations of governed machines connected through a high tie-line reactance. This approaches the case described in Section 2.2.4, but the governor gain is still less than one-tenth of the tie-line synchronizing-torque coefficient. As would be expected from the value of T'_{d12} , no mutual oscillations take place between rotors. The inherent frequency of such oscillations is 1.15 c/s, which is less than 5 times the governor oscillation frequency. Some interference might be expected, but oscillograms (b) in Fig. 11 and (b) in Fig. 12 show records of approximately the same frequency and damping as those of (c) in Fig. 10, the time-constant of the envelope decay in both cases being approximately 12 sec. A tie-line reactance of 1.9 p.u., therefore, does not make the governing less stable. The steady-state conditions are described in eqn. (15).

(4) CONCLUSIONS

The purpose of the paper has been threefold. First, it has been to show that time-error governed turbines driving alternators in a synchronous system, with circuit parameters usually found in power systems, can work with stability and produce the required steady-state errors; to do this convincingly, something more than the solutions of the equations in Section 2 by means of a differential analyser is required, and a model system having the same characteristics proves it beyond doubt. Secondly, the purpose has been, in the new scheme of control proposed, to consider the new practical operating techniques required, and a model system is the best way of doing this. Thirdly, in a control system in which angle is the controlled quantity, to show the need for careful consideration to the application to tie-lines, where the angle between ends may be large and variable.

In connection with the first point, time-error governors with reasonably high gains require stabilizing feedback. For the governor action to contribute to the electrical stability of the system during the critical first swing, a very high governor gain is necessary. Such a gain appears to require stabilizing circuits giving conditional stability only; any change in gain will make the system unstable.

Of all the parameters of the electrical system, the most important from the point of view of governor operation appears to be the damping coefficient T_d . In a governed system, other than one in which the governor gain approaches the electrical synchronizing-torque coefficient, the damping comes from the prime mover and the electrical load. High values of T_d arising from the action of alternator amortisseur windings have little effect.

Apart from the value of governor gain used, the time lags associated with the throttle-valve servo-motor and the steam header have a profound effect on governor stability. The values used in the laboratory apparatus for these time-constants are about 0.3 sec; this represents modern steam practice. However, a high-head hydro-electric set might have a penstock time-constant of 4.0 sec. The effect of the stabilizing feedback is to mask these time-constants but not to remove their effect completely.

It is suggested that the results of tests given will lead to governor parameters which may be recommended for use in a time-governed power system. The conservative system of

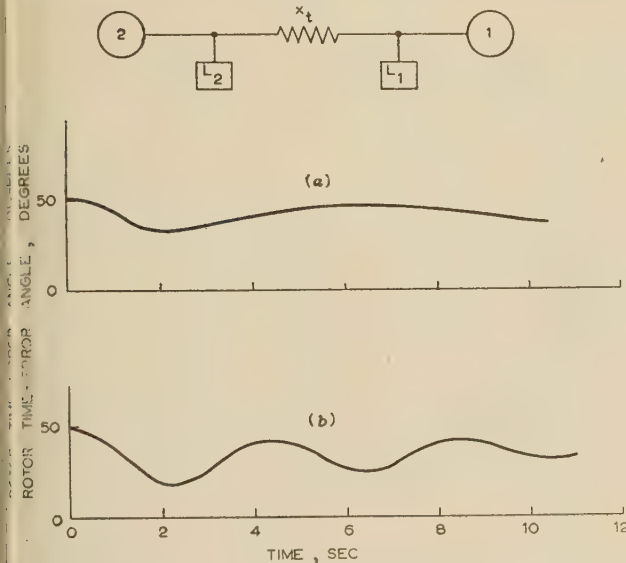


Fig. 11.—Record of $\Delta\theta_1$; Sets 1 and 2 in parallel; 0.85 and 0.93 p.u. load.

$$\begin{aligned} T_{a1} = T_{a2} &= 1050 \text{ rad.} & T'_a &= 525 \text{ rad.} \\ T_{d1} &= 3.5. & T'_{d12} &= 2.1. \\ T_{d2} &= 3.1. & T'_{d12} &= 1.2. \\ K_1 &= 0.08, K_2 = 0.07. & x_t &= 1.9 \text{ p.u.} \\ T'_1 = T'_2 &= 0.35 \text{ sec.} \\ \Delta M_t &= 0.042 \text{ p.u.} \\ (a) \quad T'_1 = T'_2 &= 0. \\ (b) \quad T'_1 &= 0.29 \text{ sec; } T'_2 = 0.32 \text{ sec.} \end{aligned}$$

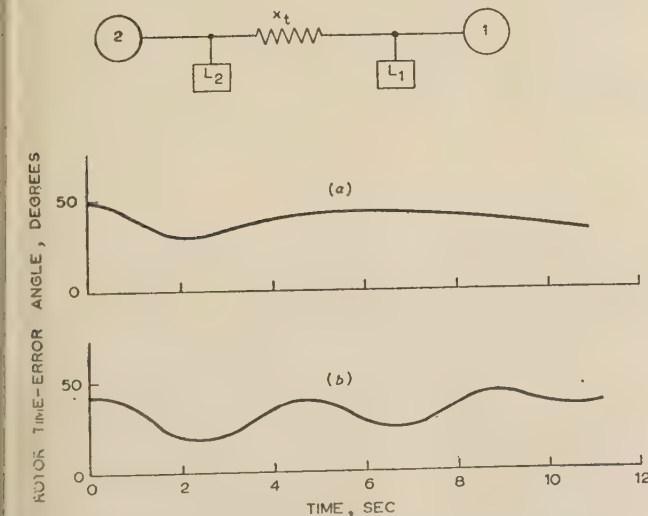


Fig. 12.—Record of $\Delta\theta_2$; Sets 1 and 2 in parallel; 0.85 p.u. and 0.93 p.u. load.

Data as in Fig. 11.

stabilizing used limits the governor gain K to less than 0.08 p.u./rad for a typical steam turbine having steam-header and valve servo-motor time-constants of 0.3 sec (approximately) and 0.35 sec, respectively. The total load and prime-mover damping used (T_d) ranges from 2.5 to 3.5; it must be borne in mind that it is not always possible to rely on as much damping as this, in which case the governor gain must be reduced if the system is to remain stable. On no-load the value of T_d is very small and the governor gain must again be correspondingly reduced.

The test oscillograms show the response of the governed system to be somewhat oscillatory. There is a tradition among speed-governor designers requiring them to make their systems critically damped or over-damped; however, there does not appear to be any reason why appreciable overshoot should not be allowed in the interests of speedy response, and, indeed, alternator rotors can oscillate with respect to each other for a short time after a switching disturbance with little ill effect. This is exactly what was happening in the tests; although it is not suggested that the degree of stability of, e.g., oscillograms (c) Fig. 10, (b) Fig. 11 and (b) Fig. 12 is sufficient, the point is made that a critically damped system is not necessary or desirable.

In the practical design of the governor attention must be paid to the way in which the throttle setting control can be used to allow the error pick-off synchro to work in the middle of its characteristic. The operating value of governor gain in conjunction with the maximum operating angle swing of the synchro (about 60°) will determine the magnitude of the load impact or swing that the set can take. The effective angle swing of the synchro can be increased by gearing, but this would mean that any drift in the synchro supply voltage would be accentuated in its effect on system load level.

The throttle setting control (which is a bias voltage in the servo-motor amplifier, or alternatively a device for rotating the stator of the governor synchro) may also be used to take up large angular differences between two systems connected by a long tie-line. It is through the throttle setting voltage that the tie-line controllers and stabilizers operate in their action on the turbine input torque. Apart from the necessity of taking up the tie-line operating angle, time-error governed areas connected by a long tie-line work with much the same degree of stability as closely coupled areas.

(5) ACKNOWLEDGMENTS

No work of any practical value has ever been accomplished without the help of many people. This project has been no

exception, and I am deeply indebted to those on the staff of Melbourne University and elsewhere for assistance so readily given. Professor C. E. Moorhouse has been characteristically generous in making available time and facilities without which little could have been achieved.

(6) REFERENCES

- (1) BROADBENT, D.: 'Power System Frequency, Tie-Line and Time Control' (M.Eng.Sc. Thesis, Melbourne University, 1953).
- (2) BROADBENT, D.: 'Integral Governing of Turbo-Alternators', *Electrical Engineer* (Melbourne), 1953, **29**, p. 354.
- (3) BROADBENT, D.: 'Stability of Integral Governing', *ibid.*, 1955, **32**, p. 40.
- (4) BROADBENT, D.: 'The Stability of Time-Error Governed Turbines in Power Systems', *Australian Journal of Applied Science*, 1955, **6**, p. 281.
- (5) CAHEN, F., and CHEVALLIER, A.: 'Le réglage puissance-phase. Nouvelle méthode pour le réglage automatique de la fréquence d'un réseau comportant de multiples usines génératrices', *Bulletin de la Société Française des Électriciens*, 1953, **3**, (VII), No. 34.
- (6) CAHEN, F., and CHEVALLIER, A.: 'Automatic Frequency Regulation in Large Networks', C.I.G.R.É., Paris, 1954, Paper No. 339.
- (7) CAHEN, F.: 'Load-Phase Tie-Line Energy Control of Inter-connected Power Systems', *Transactions of the American I.E.E.*, 1955, **74**, Part III, p. 1.
- (8) CAHEN, F.: 'The Problem of Frequency and Power Transfer Regulation', UNIPED Congress, London, 1955, Paper No. III, 1.
- (9) CONCORDIA, C., and KIRCHMAYER, L. K.: 'Tie-Line Power and Frequency Control of Electric Power Systems', *Transactions of the American I.E.E.*, 1953, **72**, Part III, p. 562.
- (10) CONCORDIA, C., and KIRCHMAYER, L. K.: 'Tie Line Power and Frequency Control—II', *ibid.*, 1954, **73**, Part III, p. 133.
- (11) CURTIS, T. A.: 'Power System Governing', *Electrical Engineering*, 1947, **56**, p. 1197.
- (12) CURTIS, T. A., and PICKSLAY, W. M.: 'Preliminary Consideration of System-wide Load Division by means of Accurate Time Control', *Transactions of the American I.E.E.*, 1948, **67**, p. 1562.

RADIO PROPAGATION OVER A DISCONTINUITY IN THE EARTH'S ELECTRICAL PROPERTIES—II. COASTAL REFRACTION

By T. B. A. SENIOR, M.Sc., Ph.D.

(The paper was first received 20th May, 1955, and in revised form 14th June, 1956. It was published as an INSTITUTION MONOGRAPH in October, 1956.)

SUMMARY

The paper presents merely one aspect of the general problem of propagation over a coastline, but special consideration is given to it since it differs from other effects in being present only at oblique incidence. By adopting a model for the coastal region the analysis given in Part I of the paper (see page 43) may be used to provide expressions for the angle of refraction appropriate to various positions of the transmitter and receiver relative to the coast. It is shown that the onset of the refraction occurs abruptly at the coastline, the theory suggesting that the direction of propagation is initially turned through an angle $\pi/2$. Within a wavelength of the coast, however, the refraction depends critically upon the way in which it is defined, and expressions which are equivalent at larger distances here lead to widely differing values. Nevertheless, if the transmitter is far away, it is certain that the refraction is largely confined to within a few wavelengths of the discontinuity and, moreover, decreases as the receiver recedes from the coastline; ultimately the original direction of propagation is restored. But, if the transmitter is itself within the coastal region, there is a permanent distortion of the field and a directional error which is virtually independent of the position of the receiver.

(1) INTRODUCTION

When a radio wave crosses a discontinuity in the earth's electrical properties its direction of propagation may be appreciably changed at points close to the discontinuity. This phenomenon is usually termed 'coastal refraction' and was first observed during the 1914-18 War.

It was natural to seek an explanation in terms of the difference in the velocities of propagation on the two sides of the boundary, for it was known experimentally that the velocity increases with increasing conductivity. Although the change in velocity is small, it was thought to be sufficient to affect the direction of a wave passing over, for example, a coastline.

The first treatment was by T. L. Eckersley,¹ who reasoned by analogy with ordinary refraction theory. At a surface separating two distinct media the angle through which an incident ray is deviated may be shown to be

$$\Omega = \frac{n^2 - 1}{2n} \tan \beta$$

where β is the angle of incidence and n is the refractive index of one medium relative to the other. By interpreting n as the ratio of the velocity, v_L , over land to that, v_s , over sea and using the theory of Zenneck wave propagation, Eckersley was led to a refraction of the correct sign (i.e. towards the normal on crossing from sea to land), but of a magnitude much less than that observed in practice.

The Zenneck wave, however, travels at a greater effective velocity the lower the conductivity of the earth, a fact which may be verified from the tilt of the wavefronts. Eckersley had failed

to realize this and had accidentally inverted the ratio v_s/v_L ; the agreement with the observed sign of refraction was thus removed.

Further experimental measurements of coastal refraction were given by Smith-Rose,² who reported that 'the magnitude appears to be of the order 3° to 5° , although greater errors [in bearing] of up to about 10° are sometimes experienced', but no theory was put forward to account for them. Indeed, the failure of the Zenneck wave analysis led Barfield³ to conclude that the refraction must be the result of causes other than the difference in superficial velocity. He suggested that an alternative explanation lay in the continued transference of energy across the boundary at right angles to the general direction of travel of the waves; this would give rise to a kink in the wavefronts near to the boundary and would explain the facts qualitatively.

For a further 17 years, however, coastal refraction remained a largely unsolved problem, and it was not until 1942 that a satisfactory theory was developed. The work of Grünberg^{4,5} and Feinberg⁶⁻⁸ has been discussed in Part I,⁹ but it should be emphasized that they were concerned with a study of coastal refraction in addition to mixed-path attenuation. For an incident plane wave Grünberg obtained a refraction of the correct sign and showed that it decreased with increasing distance from the coast. The theory was generalized by Feinberg⁸ to the case of a point-source transmitter located at a finite distance from the boundary, and expressions were given for the refraction at various positions of the receiver relative to the coast.

More detailed experimental evidence of the refraction was also available as a result of the work of Alpert and Gorozhankin,¹⁰ who carried out an investigation of the field structure at a site on the Black Sea coast. Curves of equal phase velocity and bearing error were plotted, and these reveal a very complicated type of field behaviour. The main peculiarity is a decrease of phase velocity over the sea in directions close to $\beta = \pi/4$ at a distance of two to five wavelengths. With increasing distance from the coast the complex field structure disappears, and the phase velocity approaches that of free space. As regards radio direction-finding from a transmitter on land, the bearing error changes sign at $\beta = \pi/4$, a behaviour which Feinberg⁸ has sought to explain by considering a transition region separating the homogeneous land and sea media, and by placing the 'equivalent shore line' some distance behind the actual coast. Similar results were found for a transmitter over the sea.

Meanwhile, the use of war-time radio-navigation systems had provided a more precise knowledge of the velocity at which waves are propagated over the surface of a relatively homogeneous earth. It was known that the Zenneck theory was irrelevant as regards propagation from an aerial near the ground, the true 'surface wave' being of the kind calculated by Norton.^{11,12} Near to the source this wave undergoes a progressive change of phase, corresponding to a velocity less than that in free space, and the velocity becomes equal to that in free space only at large distances. The variation of velocity with distance is in agreement with that obtained by Alpert *et al.*¹³ and Ryazin¹⁴ using methods different from that of Norton.

Correspondence on Monographs is invited for consideration with a view to publication.
T. B. A. Senior, who was formerly at Cambridge University, is now at the Radar Research Establishment.

The fact that the velocity (in practice) is less than that in free space has been amply confirmed by experiment (see, for example, Ratcliffe¹⁵), and led T. L. Eckersley¹⁶ to resurrect his original method for treating coastal refraction. With the Sommerfeld, rather than the Zenneck, theory of propagation, he succeeded in deriving an angle of refraction of the correct sign, but the solution appears to indicate that there is no refraction on short or very long waves. Moreover, the refraction is of the required magnitude only if there is a 2% difference in the velocities over land and sea; such a difference cannot be accounted for on the basis of an ideal earth with no irregularities, but would occur 'if the land emerged from the sea in a hill of about 2000 m radius'. Indeed, the analysis suggests that, if the land is flat where it joins the sea, there should only be a negligible amount of refraction.

The reason for the failure of Eckersley's rejuvenated theory lies not in the values taken for the velocities of propagation but in the formula for the refraction in terms of these velocities. The formula is derived on the assumption that coastal refraction is analogous to refraction in the optical sense and is really an expression for the deviation of a ray incident on the interface between two media representing land and sea. The analogy, however, is unjustified, since in coastal refraction the wave never actually changes the medium in which it is travelling.

In spite of its name, coastal refraction is essentially a diffraction phenomenon (see Reference 17), and once this has been realized it seems likely that the field will be profoundly disturbed in the immediate vicinity of the coastline. Such a disturbance does occur within (about) a wavelength of the coast, and the refraction is primarily due to the rapid change of phase velocity in this region, rather than to the difference of velocities for land and sea propagation. If Eckersley's formula has any bearing on the problem at all, its application would appear to be restricted to distances from the coast at which the velocity has settled down to that characteristic of the particular medium. At such distances the refraction is negligible, the influence of the coastline having disappeared.

The present theory attacks the problem from the diffraction aspect rather than from the change in velocity at the boundary, and in this respect the approach is similar to that of the Russian authors. On the other hand, Clemmow's more rigorous discussion¹⁸ of the mixed-path problem is confined to propagation normal to the boundary and is therefore inappropriate here, although justification for applying the solution to each radial separately does exist by virtue of the remarks in Section 5 of Part I.

An idealized form of coastal region is again assumed and is replaced by the model represented in Section 2. The solution for any type of transmitter can then be obtained by using the technique described in Section 4.4 of Part I, and the resulting expressions for the field components provide the basis for the analysis contained in the present paper.

Since coastal refraction occurs only at oblique incidence, it is necessary to choose either a point-source transmitter or a three-dimensional plane wave. The latter is treated in Section 3, and it is shown that many of the accepted features of coastal refraction can be explained in terms of this simple solution. Indeed, the expression for the angle of refraction is almost identical with that derived in Section 4 for the more realistic case of a dipole source.

(2) THE MODEL FOR THE IDEALIZED PROBLEM

The problem considered is that in which vertically polarized waves are propagated across a straight-line discontinuity in the electrical properties of a smooth flat earth. The situation is further specialized by the assumption that one of the media (medium 2) has infinite conductivity, and this is replaced by an

infinitely thin, perfectly conducting half-plane lying in the interface of the air (regarded as free space) and medium 1, the latter being taken to fill completely the region below the interface (see Fig. 1). Such a model, of course, is a valid representation of the



Fig. 1.—The model of the idealized problem.

physical problem only as regards the determination of the field above the earth's surface, and attention will be confined to this region.

The assumption of perfect conductivity for an earth constituent may sometimes be justified, sea water, for example, often fulfilling the condition to an adequate degree of accuracy. For this reason media 1 and 2 will be referred to as land and sea respectively, but whilst consideration will mainly be given to an investigation of the behaviour of the field near to a coastline, the results obtained are appropriate to any mixed-path problem in which the above assumptions may be supposed satisfied.

(3) THE TRANSMITTER AT INFINITY

If the transmitter is sufficiently far from the coastline, the incident field may be treated as a plane wave. The simplicity of the solution in this case makes it an ideal one with which to commence a study of coastal refraction.

(3.1) The Dependence on Field Component

As in Part I the perfectly conducting sheet is assumed to occupy the half-plane $y = 0, x > 0$ of some Cartesian co-ordinate system (x, y, z) , the remainder of the plane $y = 0$ representing the surface of a land of complex refractive index $1/\eta$. A 3-dimensional plane wave is incident in the direction $(-\cos \alpha \cos \beta, -\sin \alpha \cos \beta, \sin \beta)$ and is defined in terms of the wave function*

$$w = e^{j\gamma(x \cos \alpha \cos \beta + y \sin \alpha \cos \beta - z \sin \beta)}$$

by eqns. (29) in Part I; it is characterized by the absence of any component of magnetic vector normal to the earth, and the appropriate Fresnel reflection coefficient for a land of infinite extent is therefore

$$\rho = \frac{\sin \alpha \cos \beta - \eta}{\sin \alpha \cos \beta + \eta}$$

The total field was obtained in Part I, Section 3.2, by an elementary method consisting of a combination of ray theory and diffraction theory. The solution is valid providing the transmitter lies far over the land, and if the receiver is on the surface of the sea the only non-zero field components are specified by

$$A(E_y) = 2 \left(1 - (1 - \rho) \frac{e^{j\pi/4}}{\sqrt{\pi}} \left\{ F[\sqrt{(2\gamma x \cos \beta)} \cos \tfrac{1}{2}\alpha] + j \cos \tfrac{1}{2}\alpha \frac{\sin^2 \beta + \cos^2 \beta \cos \alpha}{1 - \sin^2 \alpha \cos^2 \beta} \frac{e^{-j\gamma x(1 + \cos \alpha) \cos \beta}}{\sqrt{(2\gamma x \cos \beta)}} \right\} \right)$$

* Rationalized m.k.s. units are employed and a time factor $e^{j\omega t}$ is suppressed throughout; γ is the propagation coefficient.

$$A(H_x) = 2 \left\{ \left[1 - (1 - \rho) \frac{e^{j\pi/4}}{\sqrt{\pi}} \left\{ F[\sqrt{(2\gamma x \cos \beta) \cos \frac{1}{2}\alpha}] + j \cos \frac{1}{2}\alpha \frac{e^{-j\gamma x(1 + \cos \alpha) \cos \beta}}{\sqrt{(2\gamma x \cos \beta)}} \right\} \right] \right\}$$

$$A(H_z) = 2 \left\{ 1 - (1 - \rho) \frac{e^{j\pi/4}}{\sqrt{\pi}} F[\sqrt{(2\gamma x \cos \beta) \cos \frac{1}{2}\alpha}] \right\}$$

where the symbol A denotes the factor by which the free-space field must be multiplied to give the total field, and $F(a)$ is the Fresnel integral

$$F(a) = \int_a^\infty e^{-js^2} ds$$

Of all components these three are least affected by the approximate nature of the solution, suggesting that an enhanced accuracy can be achieved by basing the subsequent analysis upon them. Such a policy has the added virtue of being convenient from the practical standpoint, since it is usually the vertical electric or the horizontal magnetic field that is measured.

In the region where coastal refraction is important the argument of the Fresnel integral is small [for example, when $\alpha = 170^\circ$, $\sqrt{(2\gamma x \cos \beta) \cos \frac{1}{2}\alpha} < 0.1$ if $x \cos \beta < 16.5\lambda$] and this allows the integral to be replaced by its series expression to give

$$A(E_y) = \frac{2 \sin \alpha \cos \beta}{\eta + \sin \alpha \cos \beta} \left[1 + \eta e^{j\pi/4} \operatorname{cosec} \frac{1}{2}\alpha \sqrt{\left(\frac{2\gamma x \sec \beta}{\pi} \right)} \left(1 - j \frac{\sin^2 \beta + \cos^2 \beta \cos \alpha}{1 - \sin^2 \alpha \cos^2 \beta} \frac{e^{-2j\gamma x \cos \beta \cos^2 \frac{1}{2}\alpha}}{2\gamma x \cos \beta} \right) \right] \quad (1)$$

$$A(H_x) = \frac{2 \sin \alpha \cos \beta}{\eta + \sin \alpha \cos \beta} \left[1 + \eta e^{j\pi/4} \operatorname{cosec} \frac{1}{2}\alpha \sqrt{\left(\frac{2\gamma x \sec \beta}{\pi} \right)} \left(1 - j \frac{e^{-2j\gamma x \cos \beta \cos^2 \frac{1}{2}\alpha}}{2\gamma x \cos \beta} \right) \right] \quad (2)$$

$$A(H_z) = \frac{2 \sin \alpha \cos \beta}{\eta + \sin \alpha \cos \beta} \left[1 + \eta e^{j\pi/4} \operatorname{cosec} \frac{1}{2}\alpha \sqrt{\left(\frac{2\gamma x \sec \beta}{\pi} \right)} \right] \quad (3)$$

The dependence of the factor A upon field component is essential in order that the conditions at the edge of the conducting sheet be satisfied. Thus, E_y and H_x become infinite as $x \rightarrow 0$ whilst H_z remains finite, and these contrasting types of behaviour are reflected in the presence or otherwise of the terms in $1/x$. Such terms, however, are only appreciable near to the coast. Taking first the component E_y , the term in $1/x$ may be neglected if

$$\gamma x \gg \frac{1}{2} \sec \beta |\cos 2\beta| \quad (4)$$

($\cos \alpha$ having been replaced by -1) and the corresponding relation for H_x is

$$\gamma x \gg \frac{1}{2} \sec \beta \quad (5)$$

Some values of $\frac{1}{2} \sec \beta |\cos 2\beta|$ and $\frac{1}{2} \sec \beta$ are given in Table 1.

Table 1

β	0°	15°	30°	45°	60°	75°
$\frac{1}{2} \sec \beta \cos 2\beta $	0.500	0.448	0.289	0.000	0.500	1.673
$\frac{1}{2} \sec \beta$	0.500	0.518	0.577	0.707	1.000	1.932

It will be seen that the factors A differ only within (about) a wavelength of the coast and may be assumed to be identical in calculations of mixed-path attenuation where the effect of the discontinuity extends for many hundreds of wavelengths. On the other hand, even slight differences must be taken into account in a full study of coastal refraction. The effect is here confined to within a few wavelengths of the coast, and undoubtedly a large proportion of the refraction takes place in the first wavelength.

It is natural to ask if there is any justification for applying the above expressions for the field components at distances which are not in accordance with the inequalities (4) and (5). The solution which has been obtained is not exact and does not completely satisfy the boundary conditions on that part of the land surface adjacent to the discontinuity. As a result of the mismatch it will contain a perturbation which may propagate for some distance over the sea in spite of fulfilling the boundary conditions there and having the correct behaviour at the coastline. Since coastal refraction is determined by the derivative of the phase, a small discrepancy in this could have an appreciable effect upon the angle of refraction. In Part I, however, the phase of a located transmitter was derived from the solution for a plane wave and was shown to be in remarkably close agreement with the more exact calculations of, for example, Clemmow.¹⁸ This fact gives some confidence in assuming the above solution to be valid right up to the coastline, but the main justification for so doing is based upon physical sense rather than mathematical reasoning.

To bring out the dependence on field component, the angle of refraction will be determined using each of E_y , H_x and H_z , and it is clear that the results will differ amongst themselves. There is then the problem of deciding which set (if any) has practical significance.

The obvious method for attempting a measurement is to rotate a frame aerial about a vertical axis until a minimum or null signal is received. If the field is a plane wave, a null corresponds to the loop lying along the phase front of E_y , a position which also specifies the direction of the horizontal magnetic field. With the present field structure, however, the horizontal magnetic field is elliptically polarized and there is no position of the loop for which the induced signal is a minimum independently of time. Indeed, T. L. Eckersley¹ has suggested that it is necessary to be 'several wavelengths' from the coast (to allow the wavefronts to settle down) before the angle of refraction can be definitely measured. The field components will then have the same phase, but the refraction itself will be extremely small.

In spite of these difficulties it would appear that a significance can be attached to a theoretical treatment in terms of either the vertical electric or the horizontal magnetic field. Four expressions for the angle of refraction can be obtained in this way, three of them by employing the deviation of phase fronts from their undisturbed (free-space) positions, and a formula will now be derived which enables the refraction to be deduced from the phase of any one component.

(3.2) A Formula for the Angle of Refraction

The incident field is a plane wave, and the horizontal sections of the undisturbed phase fronts are therefore straight lines perpendicular to the direction of propagation. Let $P(x, 0, z)$ be a point on the far side of the coastline at which the refraction is to be calculated, the origin of co-ordinates being at the place where the normal to the undisturbed phase front at P intersects the coast. If it is assumed that the phase is a decreasing function of distance (implying that increase of distance corresponds to a negative phase change), the undisturbed phase at P may be taken to be $\gamma(x \cos \alpha \cos \beta - z \sin \beta)$. The phase at an adjacent point Q having co-ordinates $(x + \delta x, 0, z + \delta z)$ is then

$\gamma[(x + \delta x) \cos \alpha \cos \beta - (z + \delta z) \sin \beta]$, and if Q lies on the same phase front as P,

$$\delta z = \delta x \cos \alpha \cot \beta$$

Although the disturbed phase front at P is not strictly a straight line, each small portion of it can be treated as such. Let Q be that point on the same front as P for which $\angle Q'QP = \pi/2$. If ϕ is the total phase, the difference in phase between Q' and Q is $\phi(P) - \phi(Q)$ and hence

$$\begin{aligned} Q'Q &= -\frac{1}{\gamma}[\phi(P) - \phi(Q)] \\ &\simeq \frac{1}{\gamma}\left(\frac{\partial \phi}{\partial x} \delta x + \frac{\partial \phi}{\partial z} \delta z\right) \end{aligned}$$

and

$$PQ = \sqrt{(\delta x)^2 + (\delta z)^2}$$

leading to the following expression for the angle Ω between the disturbed and undisturbed phase fronts at P:

$$\Omega = \arctan \left[\frac{1}{\gamma \sqrt{(1 + \cos^2 \alpha \cot^2 \beta)}} \left(\frac{\partial \phi}{\partial x} + \cos \alpha \cot \beta \frac{\partial \phi}{\partial z} \right) \right] \quad (6)$$

It will be observed that substitution of the free-space phase into eqn. (6) yields no contribution to Ω . This result is physically necessary and enables us to replace ϕ by ϕ_A , where this denotes the phase of the appropriate factor A. But A, and therefore ϕ_A , is independent of z. Since $\cos \alpha \simeq -1$, eqn. (6) reduces to

$$\Omega = \arctan \left(\frac{\sin \beta}{\gamma} \frac{\partial \phi_A}{\partial x} \right) \quad (7)$$

This will be regarded as the formula for the angle of refraction based upon the displacement of the phase fronts of a field component whose phase (relative to that of the free-space field) is ϕ_A . A positive value of Ω corresponds to refraction away from the normal to the coastline and in the expected direction for propagation from land to sea (that is, from a medium of low to a medium of high complex refractive index).

Although η is, in general, complex, the subsequent analysis will be carried out under the assumption that η is real (a pure dielectric earth). The expressions for Ω are only applicable to this case, but are sufficient to illustrate the type of refraction produced by each of the various field components.

(3.3) Refraction in Terms of the Component E_y

If α is replaced by π , eqn. (1) gives

$$\phi_A = \frac{\pi}{4} - \arctan \left[\frac{1 - \frac{\eta}{\sqrt{(\pi \gamma x \sec \beta)}} \frac{\cos 2\beta}{\cos \beta}}{1 + 2\eta \sqrt{\left(\frac{\gamma x}{\pi \cos \beta}\right)}} \right] \quad (8)$$

and hence, by substitution into eqn. (7),

$$\Omega = \arctan \left\{ \frac{\eta \tan \beta}{\sqrt{\pi \gamma x \sec \beta}} \frac{1 - \frac{2\eta}{\sqrt{(\pi \gamma x \sec \beta)}} \frac{\cos 2\beta}{\cos \beta} - \frac{1}{2\gamma x} \frac{\cos 2\beta}{\cos \beta}}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma x}{\pi \cos \beta}\right)} \right]^2 + \left(1 - \frac{\eta}{\sqrt{(\pi \gamma x \sec \beta)}} \frac{\cos 2\beta}{\cos \beta} \right)^2} \right\} \quad (9)$$

To appreciate this result it is convenient to introduce a new variable

$$r' = x \sec \beta$$

where r' represents the distance of the receiver from the coastline measured in the direction of propagation, and eqn. (9) then becomes

$$\Omega = \arctan$$

$$\left\{ \frac{\eta \tan \beta}{\sqrt{(\pi \gamma r')}} \frac{1 - \frac{2\eta}{\sqrt{(\pi \gamma r')}} (1 - \tan^2 \beta) - \frac{1}{2\gamma r'} (1 - \tan^2 \beta)}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi}\right)} \right]^2 + \left[1 - \frac{\eta}{\sqrt{(\pi \gamma r')}} (1 - \tan^2 \beta) \right]^2} \right\} \quad (10)$$

If $\beta \geq \pi/4$ the numerator of the fraction is always positive, leading to a positive value of Ω . If β is less than $\pi/4$, however, the numerator changes sign when $r' = r'_c$ (say), where

$$r'_c = \frac{\lambda}{2} \left\{ \frac{\eta}{\pi} (1 - \tan^2 \beta) \left[1 + \sqrt{\left(1 + \frac{\pi}{2\eta^2} \frac{1}{1 - \tan^2 \beta} \right)} \right] \right\}^2$$

and is negative for $r' < r'_c$ but positive otherwise. Some values of r'_c for a pure dielectric earth with $\eta = 1/3$ are as follows:

$$\beta = 0^\circ, \quad r'_c = 0.135\lambda$$

$$\beta = 15^\circ, \quad r'_c = 0.123\lambda$$

$$\beta = 30^\circ, \quad r'_c = 0.071\lambda$$

For $r' < r'_c$ the refraction is towards the normal on crossing from land to sea and is of opposite sign to that which is expected. In fact, inspection of eqn. (10) shows that for very small values of r'

$$\Omega \simeq \arctan \left[-\frac{\tan 2\beta}{4\eta} \sqrt{\left(\frac{\pi}{\gamma r'}\right)} \right] \quad (11)$$

provided that $\beta \neq \pi/4$. Immediately the coastline is crossed the wave is refracted through an angle $\pi/2$ towards or away from the normal according as $\beta < \pi/4$ or $\beta > \pi/4$, respectively. If the initial refraction is negative, it rapidly decreases in magnitude and becomes positive within a distance of about a tenth of a wavelength; if the refraction is initially positive, it remains so.

(3.4) Refraction in Terms of the Component H_x

From eqn. (2) the phase of the H_x component can be written as

$$\phi_A = \frac{\pi}{4} - \arctan \left[\frac{1 + \eta \sec \beta \sqrt{\left(\frac{\pi}{\gamma x \cos \beta}\right)}}{1 + 2\eta \sqrt{\left(\frac{\gamma x \sec \beta}{\pi}\right)}} \right] \quad (12)$$

and the corresponding expression for Ω is

$$\Omega = \arctan \left\{ \frac{\eta \tan \beta}{\sqrt{(\pi \gamma r')}} \frac{1 + 2\eta \frac{\sec^2 \beta}{\sqrt{(\pi \gamma r')}} + \frac{1}{2} \frac{\sec^2 \beta}{\gamma r'}}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi}\right)} \right]^2 + \left[1 + \eta \frac{\sec^2 \beta}{\sqrt{(\pi \gamma r')}} \right]^2} \right\} \quad (13)$$

This is always positive and reduces to the same form as (10) when $\gamma r'$ is much greater than unity. If β is sufficiently large, say $\beta > 60^\circ$, eqn. (13) is essentially equivalent to eqn. (10) for all values of r' , the approximation in going from eqn. (10) to eqn. (13) being to neglect unity by comparison with $\tan^2 \beta$. For small β , however, the differences between the values of Ω

determined from the E_y and H_x components may be appreciable, particularly at points very near to the coastline.

(3.5) Refraction in Terms of the Component H_x

From eqns. (3) and (7),

$$\Omega = \arctan \left\{ \frac{\eta \tan \beta}{\sqrt{(\pi \gamma r')}} \frac{1}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi} \right)^2 + 1} \right]} \right\} \quad (14)$$

and this also is never negative. It is, in fact, the limiting form of eqns. (10) and (13) when $\gamma r'$ is large compared with unity and is identical with eqn. (10) when $\beta = \pi/4$.

(3.6) Refraction in Terms of the Horizontal Magnetic Field

If the total field were a plane wave the direction of the horizontal magnetic field relative to the x -axis would be given by

$$\tan \Gamma = |H_z|/|H_x|$$

and although the assumption of a plane wave is valid in the present problem only at large distances from the coastline, it is clear that some measure of the refraction is provided by the complement of the angle Γ . Since

$$H_z = Y \cos \alpha \cos^2 \beta \epsilon^{j\gamma(x \cos \alpha \cos \beta - z \sin \beta)} A(H_z)$$

$$\text{and } H_x = Y \sin \beta \cos \beta \epsilon^{j\gamma(x \cos \alpha \cos \beta - z \sin \beta)} A(H_x)$$

$$\text{it follows that } \frac{H_z}{H_x} = \cos \alpha \cot \beta \frac{A(H_z)}{A(H_x)}$$

and hence, from eqns. (2) and (3),

$$\tan \Gamma = \cot \beta \sqrt{\frac{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi} \right)^2 + 1} \right]}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi} \right)^2 + 1} + \left[1 + \frac{\eta \sec^2 \beta}{\sqrt{(\pi \gamma r')}} \right]^2 \right]}}$$

For the undisturbed field $\Gamma = \Gamma^i$, where $\tan \Gamma^i = \cot \beta$, and if the angle of refraction is taken to be $\Omega = \Gamma^i - \Gamma$, then

$$\Omega = \beta - \arctan \left\{ \tan \beta \sqrt{\frac{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi} \right)^2 + 1} \right]}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma r'}{\pi} \right)^2 + 1} + \left[1 + \frac{\eta \sec^2 \beta}{\sqrt{(\pi \gamma r')}} \right]^2 \right]}} \right\} \quad (15)$$

which is obviously never negative. The expression, however, is not easily compared with those obtained in Sections 3.3–3.5, and it is more convenient to rely on a comparison of numerical values.

(3.7) A Discussion of the Results

The results shown in Table 2 exhibit several striking features. For small values of β all the four methods lead to approximately

Table 2

r'	Ω^*										
	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
$\beta = 15^\circ$											
(a)	18.0	12.0	6.5	4.5	3.5	2.5	2.0	2.0	1.5	1.5	1.5
(b)	34.0	17.5	8.5	5.5	4.0	3.0	2.5	2.0	2.0	1.5	1.5
(c)	26.0	15.0	8.0	5.0	4.0	3.0	2.5	2.0	2.0	1.5	1.5
(d)	27.0	15.0	8.0	5.0	4.0	3.0	2.5	2.0	2.0	1.5	1.5
$\beta = 30^\circ$											
(a)	43.5	27.0	15.0	10.0	7.5	6.0	5.0	4.0	3.5	3.0	2.5
(b)	77.0	39.0	19.0	12.0	9.0	7.0	5.5	4.5	4.0	3.5	3.0
(c)	56.0	31.0	16.0	11.0	8.0	6.0	5.0	4.5	4.0	3.5	3.0
(d)	58.0	32.0	17.0	11.0	8.0	6.5	5.0	4.0	3.5	3.0	2.5
$\beta = 45^\circ$											
(a)	96.0	54.0	28.0	19.0	14.0	11.0	9.0	7.5	6.5	5.5	5.0
(b)	149.0	74.0	35.0	22.5	16.0	12.5	10.0	8.5	7.0	6.0	5.5
(c)	96.0	54.0	28.0	19.0	14.0	11.0	9.0	7.5	6.5	5.5	5.0
(d)	96.0	56.0	29.0	19.0	14.0	11.0	9.0	7.5	6.5	5.5	5.0
$\beta = 60^\circ$											
(a)	259	128	61.0	39.0	28.0	21.5	17.5	14.5	12.0	10.5	9.0
(b)	326	157	72.0	45.0	32.0	24.0	19.5	16.0	13.5	11.5	10.0
(c)	167	94.0	49.0	32.0	24.0	19.0	15.0	13.0	11.0	9.5	8.5
(d)	174	99.0	52.0	34.0	25.0	19.5	16.0	13.5	11.5	10.0	9.0
$\beta = 75^\circ$											
(a)	4194	496	237	146	102	76.0	60.0	49.5	40.5	34.5	30.0
(b)	919	575	260	182	112	83.0	65.0	52.5	43.5	37.0	32.0
(c)	359	212	109	71.0	52.0	40.5	33.0	27.5	23.5	20.5	18.0
(d)	323	169	94.0	63.0	46.0	37.0	29.5	24.5	21.0	18.0	16.0

* The angle of refraction Ω (measured in minutes) for a purely dielectric earth ($\eta = \frac{1}{2}$) is calculated using

- (a) the deviation of the phase fronts of E_y (Section 3.3),
- (b) the deviation of the phase fronts of H_x (Section 3.4),
- (c) the deviation of the phase fronts of H_z (Section 3.5),
- (d) the change in direction of the horizontal magnetic field (Section 3.6); the distance r' (in the direction of propagation) is measured in wavelengths.

the same value of Ω at points more than half a wavelength from the coast. As β increases the methods (c) and (d) continue to be in agreement, but (a) gives values less than, and (b) values greater than, these for small values of r' . When $\beta = 45^\circ$, methods (a), (c) and (d) yield almost identical results. For larger values of β , (a) and (b) are in agreement, as are (c) and (d), but the former pair give results considerably in excess of the latter. This difference becomes more pronounced as β increases.

It should be remembered that r' is the distance from the coast-line measured in the direction of propagation, and for large values of β , even quite large values of r' correspond to points still in the immediate vicinity of the coast; e.g. if $\beta = 75^\circ$, $r' = 10\lambda$ is equivalent to a normal distance of only 2.6λ . Moreover, for large values of β , exceeding 60° , say, the accuracy of the solution (see Part I, Section 3.3) decreases rapidly with increasing values of β . It would therefore be unwise to attach great significance to the above results for $\beta = 75^\circ$.

No explanation can be discovered for the anomalous behaviour, very near to the coast, of Ω found by method (a). It may be due entirely to the approximate nature of the solution, for in reality there is little justification for applying it when x is less than (about) a wavelength. Nevertheless, the above calculations suggest that the solution remains valid to within a wavelength of the coast, the values of Ω being in reasonable agreement, as regards order of magnitude, with the more reliable experimental measurements.

(4) THE TRANSMITTER AT A FINITE DISTANCE FROM THE COAST

A discussion of coastal refraction based entirely upon the solution for a plane wave is only appropriate to the case in which the transmitter can be assumed to be at infinity. When the transmitter is at a finite distance, a similar treatment is possible in terms of the solution for a point source, and this will now be followed through.

(4.1) A General Expression for Ω

The transmitter is taken to be a vertical electric dipole situated over land and at a large but finite distance from the coastline. If its height y_0 is in accordance with the inequalities [Part I, eqn. (9)], the approximate technique described in Part I, Section 5, is applicable, and for a receiver on the surface of the sea the solution reduces to the form

$$E_y = \frac{2y_0/r'_0}{y_0/r'_0 + \eta} \left[1 + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma R r'}{r'_0} \right)} \right] \frac{e^{-j\gamma R}}{\gamma R} \quad (16)$$

[see eqn. (18) Part I], where r'_0 and r' are the distances of the transmitter and receiver, respectively, from the point at which the direct ray crosses the coast. The angle which this ray makes with the normal to the coastline is denoted by β .

Using eqn. (16) the modification to the free-space field is given by the factor

$$A = \left[1 + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma R r'}{r'_0} \right)} \right] \quad (17)$$

and since y_0 does not occur explicitly in this expression, all dependence on y_0 having been absorbed by the 'effective' height-gain function $(y_0/r'_0)/(y_0/r'_0 + \eta)$, it is reasonable to suppose that eqn. (17) will also be valid for a ground-level transmitter. Although it will be assumed that this is so, it is not essential for the purposes of the analysis, and, indeed, to take $y_0 \neq 0$ merely serves to complicate the results, the magnitude of the refraction being virtually unchanged in view of the restrictions upon y_0 imposed by the conditions [eqn. (9) Part I].

The angle of refraction will be calculated from the deviation

of the phase fronts of E_y , this being the only field component readily available. A formula for Ω in terms of the rate of change of phase with distance from a point source can be obtained in a manner directly analogous to that for a plane wave, and bearing in mind that the horizontal sections of the undisturbed phase fronts are now circles centred on the transmitter instead of straight lines perpendicular to the direction of free-space propagation, it can be shown that

$$\Omega = \arctan \left(-\frac{1}{\gamma R} \frac{\partial \phi_A}{\partial \beta} \right) \quad (18)$$

where ϕ_A is the phase of the factor A , and R, β are independent polar co-ordinates referred to the transmitter as origin. A similar formula has been employed by Feinberg.⁸

To determine Ω it is necessary to consider the variation of phase with angle β . Let l be the normal distance of the transmitter from the coast.

Then

$$r'_0 = l \sec \beta$$

and

$$r' = R - l \sec \beta$$

$$\text{giving } A = \left\{ 1 + \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left[2\gamma R \left(\frac{R}{l} \cos \beta - 1 \right) \right]} \right\}$$

Hence

$$\phi_A = \frac{\pi}{4} - \arctan \left\{ \frac{1}{1 + 2\eta \sqrt{\left[\gamma R \left(\frac{R}{l} \cos \beta - 1 \right) \right]}} \right\} \quad (19)$$

and since R, β are the only variable co-ordinates, eqn. (18) can now be used to provide an expression for Ω in the form

$$\Omega = \arctan \left[\frac{\tan \beta}{4\gamma r'} \frac{1}{1 + \eta \sqrt{\left(\frac{\gamma R r'}{\pi r'_0} \right)} + \frac{1}{2\eta} \sqrt{\left(\frac{\pi r'_0}{\gamma R r'} \right)}} \right] \quad (20)$$

(4.2) The Nature of the Refraction

It is clear that the above expression can never be negative, and the refraction is therefore away from the normal on crossing from land to sea. This is true even at points near to the coastline. In fact, for positions of the receiver such that $r' < \lambda/10$ (say) the term $[1 + \eta \sqrt{(\gamma R r'/\pi r'_0)}]$ can be neglected by comparison with $1/2\eta \sqrt{(\pi r'_0/\gamma R r')}$ to give

$$\Omega \simeq \arctan \left[\frac{\eta \tan \beta}{4\sqrt{(\pi \gamma r')}} \right] \quad (21)$$

from which it would appear that the refraction at the coastline is $\pi/2$ for all non-zero values of β . In contrast with this, the same method of derivation applied to the solution for an incident plane wave gave an initial refraction of $\pm \pi/2$ according as $\beta \gtrless \pi/4$.

Eqn. (20) can be used to study the nature of the refraction as the receiver moves away from the coast. Commencing with a value $\pi/2$, the angle first decreases at a rate proportional to $\sqrt{r'}$. When the approximate form of eqn. (21) is no longer applicable, a comparison of the relative magnitudes of the terms in eqn. (20) shows that the further decrease of Ω is primarily due to the factor $1/r'$. Finally, at large distances from the coast, say $r' > 40\lambda$, eqn. (20) can be simplified by neglecting $[1 + 1/2\eta \sqrt{(\pi r'_0/\gamma R r')}]$ compared with $\eta \sqrt{(\gamma R r'/\pi r'_0)}$,

$$\text{and then } \Omega \simeq \arctan \left[\frac{\tan \beta}{4\eta \gamma r'} \sqrt{\left(\frac{\pi r'_0}{\gamma R r'} \right)} \right] \quad (22)$$

which decreases as $(r')^{-3/2}$. Under most circumstances, however, Ω will have become effectively zero before eqn. (22) may be applied.

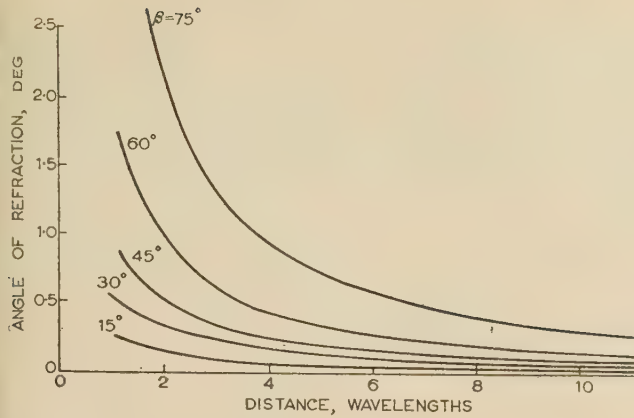


Fig. 2.—The angle of refraction for a point source as a function of the radial distance from the coastline.

In Fig. 2 the refraction Ω is plotted against r' for the five angles $\beta = 15^\circ, 30^\circ, 45^\circ, 60^\circ$ and 75° , the case considered being that of a transmitter above a pure dielectric earth ($\eta = \frac{1}{2}$) and at a normal distance of 300 wavelengths from the coast. Eqn. (20) has been used in the calculations. The rapid decrease of Ω with increasing r' is clearly shown, and for all except the largest values of β the refraction is negligible at $r' = 10\lambda$.

Comparison of the results with those obtained from the solution for an incident plane wave reveals an almost complete agreement with the calculated values based upon the deviation of the phase fronts of H_z . The relevant expressions for Ω are also similar, as may be seen by writing eqn. (20) in the form

$$\Omega = \arctan \left\{ \frac{\tan \beta}{\sqrt{(\pi \gamma r')}} \frac{\sqrt{\left(\frac{R}{r'_0}\right)}}{\left[1 + 2\eta \sqrt{\left(\frac{\gamma R r'}{\pi r'_0}\right)}\right]^2 + 1} \right\} \quad (23)$$

This is identical with eqn. (14) if $\sqrt{(r'_0/R)}$ is replaced by unity, a valid approximation for $r'_0 \gg r'$.

At first sight the agreement is a little surprising, for it would seem more natural to expect a correspondence between the two equations based upon the component E_y . We recall, however, that the solution for a point source was given in terms of E_y , the expression for which was approximated in such a way as to be analogous to that for the component H_z of the line-source solution. This has the effect of removing those features characteristic of a component E_y (one of which is the infinity at the boundary) and impresses upon E_y the form and behaviour of a component H_z . The change is immaterial at points more than a wavelength from the coast, since there $H_z \simeq Y E_y$, and in particular the rates of change of phase with distance are the same for both components, a fact which is reflected in the values of Ω shown in Table 1. But in the vicinity of the coast the components are of essentially different form, and to apply the approximation here merely converts the expression for E_y into one for H_z .

Although there is doubt as to the validity of the analysis when $x < \lambda$, a possible failure of the solution in this region is not of great practical importance. The refraction here depends critically on the manner in which it is defined, and any measurement would therefore require the isolation of one field component, a procedure which may not be feasible in view of the complicated field structure near to the coast. Furthermore, any actual coastline is likely to differ appreciably from that which has been assumed, the transition from land to sea taking place gradually, and a model in which the change-over is abrupt would then be inappro-

priate. Nevertheless, if an expression has to be obtained for the angle of refraction immediately beyond a sharp discontinuity, there is more justification for employing the approximate solution for a plane wave rather than that for a point source: in fact, it would appear that the simple solution for a plane wave provides a theory of coastal refraction of quite sufficient accuracy for most practical purposes.

One other feature of the above formulae has still to be considered, and that is the dependence of Ω upon the angle at which the field is incident. From eqn. (20),

$$\tan \Omega \propto \tan \beta$$

the constant of proportionality, κ , being a function only of r' and r'_0 . This relationship is illustrated in Fig. 3, in which the angle

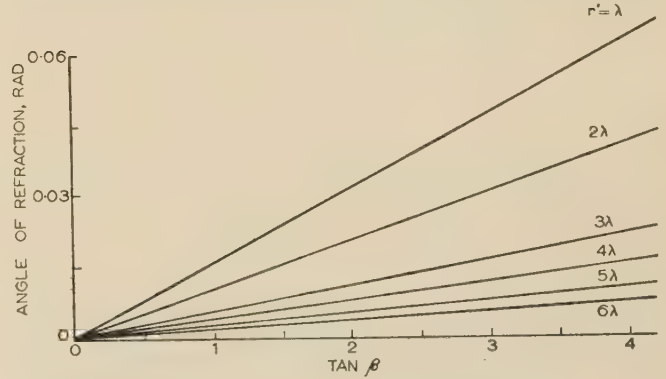


Fig. 3.—The relation between Ω and $\tan \beta$ [eqn. 20].

of refraction (in radians) is plotted against $\tan \beta$ for different values of r' . Since $\tan \Omega \simeq \Omega$, the resulting curves are straight lines passing through the origin, typical gradients being as follows:

$$r' = \lambda, \quad \kappa = 0.0157$$

$$r' = 3\lambda, \quad \kappa = 0.0055$$

$$r' = 5\lambda, \quad \kappa = 0.0031$$

Some experimental confirmation is furnished by a series of measurements of coastal refraction obtained at a direction-finding station on Cyprus, reported by T. L. Eckersley.¹ The wavelengths employed were between 800 and 1 100 m and the receiver was 'within a mile' of the coast. It was found that the angle of refraction was roughly proportional to $\tan \beta$, the gradient of the mean straight line being 0.019. In view of the variation with bearing angle of the radial distance of the receiving station from the coast, the agreement with the present theory is surprisingly good.

(5) FURTHER EXAMPLES

The preceding Sections have been devoted to a study of coastal refraction when the propagation is from land to sea. The formulae given are only applicable to this case, the restriction to a transmitter over the land being one which is imposed upon the technique for dealing with propagation over a mixed path by the model assumed for the discontinuity.

The corresponding problem of a transmitter over the sea may be treated using an alternative model in which an imperfectly conducting sheet represents the land medium. The solution for a line source on the basis of this model was outlined in Part I, Section 6, and if the modifications appropriate to a point source are carried out, we have [from eqn. (21) Part I]

$$A = \left[1 - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{\left(\frac{2\gamma r'_0 r'}{R}\right)} \right] \quad (24)$$

This is valid under the assumption that $R \simeq r'_0$, and for simplicity the factor r'_0/R will be replaced by unity. Eqn. (24) then becomes

$$A = \left\{ 1 - \frac{\varepsilon^{j\pi/4}}{\sqrt{\pi}} \eta \sqrt{[2\gamma(R - l \sec \beta)]} \right\} \quad (25)$$

leading to the following expression for the angle of refraction:

$$\Omega = \arctan \left[-\frac{\tan \beta}{4\gamma r'} \frac{1}{\eta \sqrt{\left(\frac{\gamma r'}{\pi}\right) + \frac{1}{2\eta} \sqrt{\left(\frac{\pi}{\gamma r'}\right) - 1}} \right] \quad (26)$$

In this the denominator is positive for all real values of η and r' ; the refraction is therefore negative, i.e. towards the normal, and is in the required direction for propagation from sea to land.

The above equation is similar to that given in Section 4.2, and though it will not be studied in detail, two particular cases may be noted. At points far from the coastline,

$$\Omega \simeq \arctan \left[-\frac{\tan \beta}{4\eta \gamma r'} \sqrt{\left(\frac{\pi}{\gamma r'}\right)} \right] \quad (27)$$

[cf. eqn. (22)], whilst for positions of the receiver very near to the coast (say $r' < \lambda/10$),

$$\Omega \simeq \arctan \left[-\frac{\eta \tan \beta}{2\sqrt{(\pi \gamma r')}} \right] \quad (28)$$

and comparison with eqn. (21) shows that the refraction is then exactly half that found for propagation in the reverse direction.

All the formulae so far obtained have been for a transmitter which is far from the coast and a receiver which is relatively near, but by virtue of the reciprocity condition the solutions must also be applicable to the case in which the positions of the transmitter and receiver are interchanged. As an example, consider eqn. (26). If r' and β are replaced by r'_0 and $\pi + \beta$ respectively,

$$\Omega = \arctan \left[\frac{\tan \beta}{4\gamma r'_0} \frac{1}{\eta \sqrt{\left(\frac{\gamma r'_0}{\pi}\right) + \frac{1}{2\eta} \sqrt{\left(\frac{\pi}{\gamma r'_0}\right) - 1}} \right] \quad (29)$$

and this represents the coastal refraction at a distant receiver, the transmitter being over land and near to the coast ($r'_0 \ll R$). It is independent of r' , and providing the transmitter lies in the coastal region, Ω is non-zero for all positions of the receiver, leading to a permanent directional error persisting even to infinity. This residual effect is a direct consequence of coastal refraction, and in spite of its obvious importance in connection with the siting of direction-finding stations, it does not appear to have attracted much attention. A similar result holds for a transmitter near to the coast and on the seaward side of it, as may be seen by applying reciprocity to eqn. (20).

Finally it must be emphasized that the analysis contained in the paper is based on the assumption of a pure dielectric earth, and if η is not real the formulae will be completely changed. Thus, if the frequency is sufficiently low for $\arg \eta$ to equal $\pi/4$, the equation corresponding to eqn. (26) is

$$\Omega = \arctan \left[-\frac{\tan \beta}{2\gamma r'} \frac{1}{|\eta| \sqrt{\left(\frac{2\gamma r'}{\pi}\right) + \frac{1}{|\eta|} \sqrt{\left(\frac{\pi}{2\gamma r'}\right)}} \right] \quad (30)$$

and is valid for a transmitter far out over the sea. Such frequencies have been considered by Furutsu,¹⁹ whose general expression for Ω can be reduced to manageable proportions providing the transmitter is over land and near to the coast. Application of reciprocity to his result leads to a formula which is identical to eqn. (30) when this is approximated by taking $2\gamma r' \ll \pi/|\eta|^2$.

(6) CONCLUSION

In spite of the renewed theoretical interest in coastal refraction, there is still a notable lack of reliable practical data, and most of the measurements which do exist have been obtained incidentally in the course of other work. Consequently the coastlines have not been selected especially for an investigation of refraction, and generally the influence of relief has been such as to outweigh by far any effects arising from the change of conductivity. In presenting a simplified treatment of coastal refraction it is realized that some of the theoretical assumptions are open to question, and the value of the theory for practical purposes can only be assessed by a comparison with the results of experiments more specifically designed for a study of these effects.

(7) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. P. C. Clemmow of Cambridge University for much assistance and advice, and to acknowledge the receipt of a grant from the Department of Scientific and Industrial Research for the period during which this work was carried out.

(8) REFERENCES

- (1) ECKERSLEY, T. L.: 'Refraction of Radio Waves', *Radio Review*, 1920, **1**, p. 421.
- (2) SMITH-ROSE, R. L.: 'Coastal Errors in Radio Direction Finding', *Nature*, 1925, **116**, p. 426.
- (3) BARFIELD, R. H.: 'Coastal Refraction of Radio Waves', *ibid.*, p. 498.
- (4) GRÜNBERG, G.: 'Theory of the Coastal Refraction of Electromagnetic Waves', *Journal of Physics U.S.S.R.*, 1942, **6**, p. 185.
- (5) GRÜNBERG, G.: 'Suggestions for a Theory of Coastal Refraction', *Physical Review*, 1943, **63**, p. 185.
- (6) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *Journal of Physics U.S.S.R.*, 1944, **8**, p. 317.
- (7) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *ibid.*, 1945, **9**, p. 1.
- (8) FEINBERG, E.: 'On the Propagation of Radio Waves along an Imperfect Surface', *ibid.*, 1946, **10**, p. 410.
- (9) SENIOR, T. B. A.: 'Propagation over a Discontinuity in the Earth's Electrical Properties—I' (see page 143).
- (10) ALPERT, J., and GOROZHANKIN, B.: 'Experimental Investigation of the Structure of an Electromagnetic Field over an Inhomogeneous Earth's Surface', *Journal of Physics U.S.S.R.*, 1945, **9**, p. 115.
- (11) NORTON, K. A.: 'The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, I', *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 1367.
- (12) NORTON, K. A.: 'The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, II', *ibid.*, 1937, **25**, p. 1203.
- (13) ALPERT, Y. A., MIGULIN, V. V., and RYAZIN, P. A.: 'An Investigation of the Structure of the Electromagnetic Field and the Velocity of Radio Waves', *Journal of Physics U.S.S.R.*, 1941, **4**, p. 13.
- (14) RYAZIN, P. A.: 'Propagation of Radio Waves near the Earth's Surface', *Akademiya Nauk SSR. Fizičeskii Institut Imeni P.N. Lebedeva Trudy Fizičeskogo Instituta*, 1946, **3**, p. 45.
- (15) RATCLIFFE, J. A.: 'The Velocity of Radio Waves', *Proceedings of the 7th Assembly, U.R.S.I.*, Paris, 1946.

- (16) ECKERSLEY, T. L.: Atti del Congresso Internazionale per il Cinquantenario della Scoperta Marconiana della Radio, Rome, 1947, p. 97.
- (17) BOOKER, H. G., and CLEMMOW, P. C.: 'A Relation between the Sommerfeld Theory of Radio Propagation over a Flat Earth and the Theory of Diffraction at a Straight Edge', *Proceedings I.E.E.*, Paper No. 873 R, January, 1950 (97, Part III, p. 18).
- (18) CLEMMOW, P. C.: 'Radio Propagation over a Flat Earth across a Boundary separating Two Different Media', *Philosophical Transactions of the Royal Society, A*, 1953, 246, p. 1.
- (19) FURUTSU, K.: 'Propagation of Electromagnetic Waves over a Flat Earth across a Boundary separating Different Media and Coastal Refraction', *Journal of the Radio Research Laboratories*, Tokyo, Japan, 1955, 2, p. 1.
-

STRAIN-ENERGY CALCULATIONS IN THE DESIGN OF CAT'S WHISKERS FOR SEMI-CONDUCTOR DEVICES

By S. J. MORRISON, B.Sc., A.M.I.Mech.E., Associate Member.

(The paper was first received 15th March, and in revised form 2nd June, 1956. It was published as an INSTITUTION MONOGRAPH in October, 1956.)

SUMMARY

Strain-energy methods have been used to study the mechanical behaviour of cat's-whiskers in semi-conductor devices, and the condition for zero transverse contact pressure and deflection has been expressed in the form $\int_0^L \frac{xy}{I} dl = 0$, where x and y are parameters which define the shape of the whisker and I is the moment of inertia of the whisker cross-section. For wire whiskers this is equivalent to $\int_0^L xy dl = 0$. The theory is perfectly general, and can be applied to plane whiskers of any shape. Formulae have been derived for calculating components of pressure and deflection parallel with and normal to the crystal face. Experiments on large-scale whisker models have shown close agreement with the theory.

(1) INTRODUCTION

The continuing interest in the crystal detector, after 25 years of development in radio and radar techniques, has kept the so-called 'cat's-whisker' in the limelight. In the early days of broadcasting much of the interest derived from the absorbing task of finding a 'sensitive spot' on a sparking crystal whose rugged contours ensured good mechanical stability at the point of contact. It is not generally realized that the stark smoothness of the modern crystal presents an equally absorbing problem to the designer of semi-conductor diodes, who has to ensure stability of the whisker at its point of contact on the crystal.

In semi-conductor crystal development attempts have been made to secure better electrical characteristics by using crystal surfaces of exceptional smoothness, and thus with low coefficients of friction. It is therefore a problem of practical importance to design whiskers which do not produce a transverse component of contact pressure, since transverse forces will tend to cause the whisker to skid on the crystal surface. This problem can be solved by strain-energy calculations, and the theoretical results can be tested by measurements on large-scale models. The calculations are developed from basic theorems which can be found in textbooks on strain-energy methods.*

(2) GENERAL THEORY

A hypothetical plane cat's-whisker of unspecified shape is shown in Fig. 1. The whisker is securely fixed to a rigid support at F and makes contact with a plane semi-conductor crystal surface at O , the planes of the whisker and the crystal being mutually perpendicular. It will be convenient to take co-ordinate axes X and Y through the contact point O , parallel with and normal to the crystal surface and in the plane of the whisker. The shape of the whisker can be defined in terms of the co-ordinates (x, y) of points on the whisker.

Pressure is applied by a movement of the point of contact O

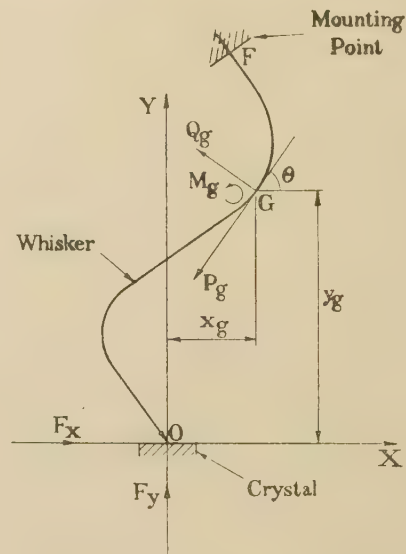


Fig. 1.—Forces acting on whisker.

relative to the mounting point F . The movement is usually normal to the crystal face, and the whisker is usually normal to the crystal at the point of contact, but these are not necessary conditions and will not be accepted as restrictions in the general development of the theory.

The shape of the contact point of the whisker (e.g. ellipsoidal, hemispherical, oblique cut, etc.) may have some effect on the limiting value of friction, but it is unlikely that any practical shape of whisker point would have the effect of introducing an angular constraint between the whisker and the crystal in the plane of the whisker. It will therefore be assumed that all types of contact point can be regarded as being equivalent to a freely pivoted joint. In a practical case this assumption will be justified if the solution leads to a transverse pressure component which is less than the limiting frictional value for the type of contact point in use.

The forces in the whisker at a general point G , at which the contour makes an angle θ with the X -axis, are the axial force, P_g , the shear force, Q_g , and the bending moment, M_g . The forces acting on the whisker at the point of contact are the transverse force, F_x , and the normal force, F_y .

By examining the equilibrium of the section OG of the whisker these forces can be related in the following way:

$$P_g = F_y \sin \theta + F_x \cos \theta \quad (1)$$

$$Q_g = F_x \sin \theta - F_y \cos \theta \quad (2)$$

$$M_g = F_y x_g - F_x y_g \quad (3)$$

It is now necessary to introduce symbols dealing with the

* PIPPARD, A. J. S.: 'Strain Energy Methods of Stress Analysis' (Longmans Green, London, 1928), pp. 19-22.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Morrison is with the British Thomson-Houston Co., Ltd.

cross-section dimensions and mechanical properties of the whisker, namely

I = Moment of inertia of whisker cross-section.

A = Area of whisker cross-section.

L = Total length of whisker contour.

E = Young's modulus.

N = Rigidity modulus.

The strain-energy expressions derived from eqns. (1)–(3) are

$$\begin{aligned} \text{Compression: } U_P &= \frac{1}{2E} \int_0^L \frac{P_z^2}{A} dl \\ &= \frac{1}{2E} \int_0^L \frac{1}{A} (F_y \sin \theta + F_x \cos \theta)^2 dl \quad (4) \end{aligned}$$

$$\begin{aligned} \text{Shear: } U_Q &= \frac{1}{2N} \int_0^L \frac{Q_z^2}{A} dl \\ &= \frac{1}{2N} \int_0^L \frac{1}{A} (F_x \sin \theta - F_y \cos \theta)^2 dl \quad (5) \end{aligned}$$

$$\begin{aligned} \text{Bending: } U_M &= \frac{1}{2E} \int_0^L \frac{M_z^2}{I} dl \\ &= \frac{1}{2E} \int_0^L \frac{1}{I} (F_y x - F_x y)^2 dl \quad (6) \end{aligned}$$

In most practical designs of whisker the strain energy due to bending will be much greater than the other two. For simplicity, the strain energies due to compression and shear will therefore be neglected, and the theory will be based solely on that due to bending. Hence:

$$\text{Total strain energy} = U \simeq \frac{1}{2E} \int_0^L \frac{1}{I} (F_y x - F_x y)^2 dl \quad (7)$$

Castigliano's first theorem can now be applied to give the deflections of the contact point C under the action of the applied forces F_x and F_y . This theorem states, 'The partial differential coefficient of total strain energy expressed in terms of the external loads, with respect to one of the external loads, is the movement of that load in its line of action.' Hence:

$$\begin{aligned} \text{Deflection } \delta_x &= \frac{\partial U}{\partial F_x} = \frac{1}{E} \int_0^L \frac{y(F_y x - F_x y)}{I} dl \\ &= \frac{F_x}{E} \int_0^L \frac{y^2}{I} dl - \frac{F_y}{E} \int_0^L \frac{xy}{I} dl \quad (8) \end{aligned}$$

$$\begin{aligned} \text{Deflection } \delta_y &= \frac{\partial U}{\partial F_y} = \frac{1}{E} \int_0^L \frac{x(F_y x - F_x y)}{I} dl \\ &= \frac{F_y}{E} \int_0^L \frac{x^2}{I} dl - \frac{F_x}{E} \int_0^L \frac{xy}{I} dl \quad (9) \end{aligned}$$

(3) CONDITIONS FOR ZERO TRANSVERSE FORCE AND DEFLECTION

The condition for zero transverse force and zero transverse deflection at the contact point are independent and may be satisfied independently or simultaneously. The three cases will be dealt with separately.

(3.1) Condition for Zero Transverse Force

This is obtained by inserting $F_x = 0$ in eqns. (8) and (9), thus

$$\delta_x = -\frac{F_y}{E} \int_0^L \frac{xy}{I} dl \quad (10)$$

$$\delta_y = \frac{F_y}{E} \int_0^L \frac{x^2}{I} dl \quad (11)$$

This can be expressed in a single equation

$$\frac{\delta_x}{\delta_y} = -\frac{\int_0^L \frac{xy}{I} dl}{\int_0^L \frac{x^2}{I} dl} \quad (12)$$

For crimped wire whiskers, in which I is constant along the entire length, the condition is

$$\frac{\delta_x}{\delta_y} = -\frac{\int_0^L xy dl}{\int_0^L x^2 dl} \quad (12a)$$

(3.2) Condition for Zero Transverse Deflection

This is obtained by inserting $\delta_x = 0$ in eqn. (8);

$$\frac{F_x}{E} \int_0^L \frac{y^2}{I} dl - \frac{F_y}{E} \int_0^L \frac{xy}{I} dl = 0$$

Therefore

$$\frac{F_x}{F_y} = \frac{\int_0^L \frac{xy}{I} dl}{\int_0^L \frac{y^2}{I} dl} \quad (13)$$

For crimped wire whiskers, in which I is constant along the entire length, the condition is

$$\frac{F_x}{F_y} = \frac{\int_0^L xy dl}{\int_0^L y^2 dl} \quad (13a)$$

(3.3) Condition for Zero Transverse Force and Zero Transverse Deflection

The simultaneous condition is given by putting $\delta_x = 0$ in eqn. (10);

$$\int_0^L \frac{xy}{I} dl = 0 \quad (14)$$

This result can also be obtained by putting $F_x = 0$ in eqn. (13). The corresponding condition for a crimped wire whisker in which I is constant along the entire length is

$$\int_0^L xy dl = 0 \quad (14a)$$

(4) PHYSICAL INTERPRETATION OF THEORY

Up to this point the theory is perfectly general and can be applied to plane whiskers of any shape. For a whisker of any specified shape it is now possible to calculate components of pressure and/or deflection, parallel and normal to the crystal face. Conversely, it is possible to specify conditions which the configuration of a practical whisker must satisfy if zero transverse force and/or deflection is to be observed.*

The following general conclusions can be drawn.

(4.1) Condition for Zero Transverse Force

In the preamble to the theory it was stated that the movement of the whisker relative to the crystal, by which contact pressure is produced, need not necessarily be normal to the crystal face. Eqns. (12) or (12a) define the direction of movement which will result in zero transverse force. Only in special circumstances (described in Section 4.3) will this direction be normal to the crystal.

(4.2) Condition for Zero Transverse Deflection

This is the condition commonly encountered in practice on devices in which contact pressure is produced by moving the whisker normal to the crystal face. Eqns. (13) or (13a) show that in the general case there may be a transverse component of contact pressure. The equations define the ratio of the normal and transverse components of pressure.

It is obvious that the transverse component of force will be negligible only if the integral $\int_0^L \frac{xy}{I} dl$ is zero, or small relative to the integral $\int_0^L \frac{y^2}{I} dl$. With in-line single-crimp whiskers of the

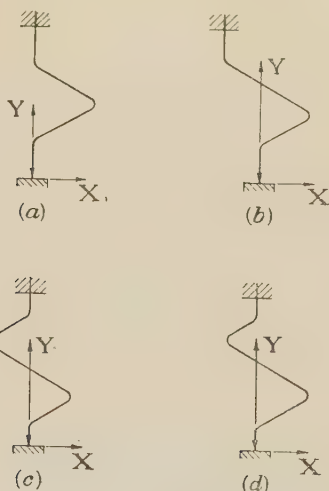


Fig. 2.—Types of crimped whisker.

- (a) In-line single crimp. (c) Symmetric double crimp.
(b) Out-of-line single crimp. (d) Corrected double crimp.

types shown in Fig. 2(a) the product xy will not change sign within the range of integration, and the integral $\int_0^L \frac{xy}{I} dl$ must therefore be finite. Consequently, single-crimp whiskers must produce a transverse component of contact pressure, and the effect will, of course, be more pronounced with short stubby whiskers.

An exception to the rule for single-crimp whiskers is the out-of-line single-crimp whisker shown in Fig. 2(b), for which the product xy will change sign where the whisker contour crosses the Y -axis. Whiskers of this type are, effectively, double-crimp whiskers and should be treated as such.

In double-crimp whiskers, such as that shown in Fig. 2(c), the product xy changes sign in the middle of the range of integration

and so the integral $\int_0^L \frac{xy}{I} dl$ can have a small value and may be zero if the whisker is suitably proportioned. Consequently, double-crimp whiskers will, in general, produce a smaller transverse component of pressure than single-crimp whiskers.

(4.3) Condition for Zero Transverse Force and Zero Transverse Deflection

The condition expressed in eqns. (14) and (14a) is the ideal one for zero transverse force on a whisker whose pressure is applied by a movement normal to the crystal face. The equations may be interpreted in the following way. For the

integral $\int_0^L \frac{xy}{I} dl$ to be zero, either x or y , or both, must change sign within the range of integration, i.e. a crimped whisker must be double-crimped, and a cantilever whisker must at some portion drop below the level of the crystal face. Furthermore, eqn. (14a) shows that in a double-crimped whisker the crimp nearest the crystal should have larger x -dimensions to compensate for the larger y -dimensions of the second crimp, as shown in Fig. 2(d). Eqn. (14) expresses a similar condition for cantilever whiskers, in which the cross-section of the whisker is also involved.

The ideal condition should not be difficult to achieve in practice in either the crimped or the cantilever types of whisker.

(5) DEFLECTION FORMULAE

The conditions for zero transverse force and/or zero transverse deflection having been defined in eqns. (12)–(14), the basic deflection formulae (8) and (9) can be modified to suit each special case.

In devices in which pressure is applied by a movement of the whisker normal to the crystal, $\delta_x = 0$ and the relation between F_x and F_y is defined in eqn. (13). In this case

$$\delta_y = \frac{F_y}{E} \left[\int_0^L \frac{x^2}{I} dl - \frac{\left(\int_0^L \frac{xy}{I} dl \right)^2}{\int_0^L \frac{y^2}{I} dl} \right] \quad (15)$$

For the ideal design of whisker, which satisfies the eqn. (14), $F_x = 0$ and $\delta_x = 0$, and the deflection formula becomes

$$\delta_y = \frac{F_y}{E} \int_0^L \frac{x^2}{I} dl \quad (16)$$

When the formulae are applied to crimped wire whiskers on which I is constant along the whole whisker length they are modified as follows:

$$\text{General case: } \delta_x = \frac{F_x}{EI} \int_0^L y^2 dl - \frac{F_y}{EI} \int_0^L xy dl \quad (8a)$$

$$\delta_y = \frac{F_y}{EI} \int_0^L x^2 dl - \frac{F_x}{EI} \int_0^L xy dl \quad (9a)$$

* British Patent Application No. 18702, 1955.

Zero transverse deflection:

$$\delta_y = \frac{F_y}{EI} \left[\int_0^L x^2 dl - \frac{\left(\int_0^L xy dl \right)^2}{\int_0^L y^2 dl} \right] \quad (15a)$$

Ideal case:
$$\delta_y = \frac{F_y}{EI} \int_0^L x^2 dl \quad (16a)$$

(6) EVALUATION OF INTEGRALS

Methods have been devised to simplify the evaluation of the definite integrals in actual design problems. With crimped wire whiskers the integrals to be calculated are

$$\int_0^L x^2 dl, \int_0^L y^2 dl \text{ and } \int_0^L xy dl$$

Practical whisker shapes can be treated as a series of straight lines and circular arcs, and the integrals can therefore be taken in stages:

$$\int_0^L f(x, y) dl = \int_0^{l_1} f(x, y) dl + \int_{l_1}^{l_2} f(x, y) dl + \int_{l_2}^{l_3} f(x, y) dl + \dots \quad (17)$$

where l_1, l_2, l_3 , etc., are the lengths of whisker contour from the contact point to points of tangency and inflection. All that is required, therefore, are formulae for evaluating the integrals for a straight line and for a circular arc at a general position in the co-ordinate field.

The general case of a straight line is illustrated in Fig. 3(a). The line of length λ connecting the points M and N makes an angle ψ with the X-axis.

Freedom equations for the line are

$$x = x_m + \lambda \cos \psi \quad (18)$$

$$y = y_m + \lambda \sin \psi \quad (19)$$

From these equations the integrals can be evaluated as follows:

$$\int_{l_m}^{l_n} x^2 dl = \lambda^3 \left[\left(\frac{x_m}{\lambda} \right)^2 + \left(\frac{x_m}{\lambda} \right) \cos \psi + \frac{1}{3} \cos^2 \psi \right] \quad (20)$$

$$\int_{l_m}^{l_n} y^2 dl = \lambda^3 \left[\left(\frac{y_m}{\lambda} \right)^2 + \left(\frac{y_m}{\lambda} \right) \sin \psi + \frac{1}{3} \sin^2 \psi \right] \quad (21)$$

$$\int_{l_m}^{l_n} xy dl = \lambda^3 \left(\frac{x_m y_m}{\lambda^2} + \frac{x_m \sin \psi + y_m \cos \psi}{2\lambda} + \frac{\sin \psi \cos \psi}{3} \right) \quad (22)$$

The general case of a circular arc is illustrated in Fig. 3(b). The arc PQ has a centre of curvature C and radius ρ and the terminal radii CP and CQ makes angles ϕ_p and ϕ_q with the X-axis.

Freedom equations of the arc are

$$x = x_c + \rho \cos \phi \quad (23)$$

$$y = y_c + \rho \sin \phi \quad (24)$$

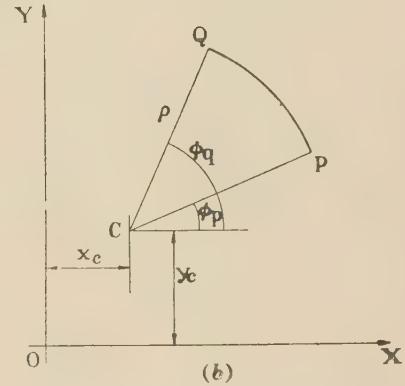
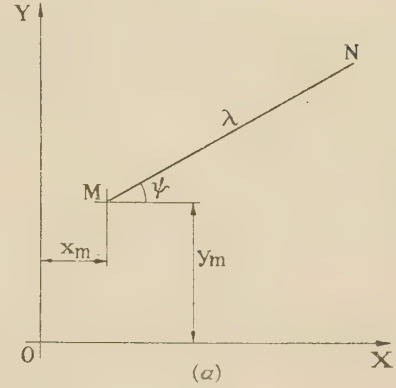


Fig. 3.—Evaluation of integrals.

(a) Straight line.
(b) Circular arc.

From these equations the integrals can be evaluated as follows:

$$\int_{l_p}^{l_q} x^2 dl = \rho^3 \left[\left(\frac{x_c}{\rho} \right)^2 + \frac{1}{2} \right] (\phi_q - \phi_p) + 2 \frac{x_c}{\rho} (\sin \phi_q - \sin \phi_p) + \frac{1}{4} (\sin 2\phi_q - \sin 2\phi_p) \quad (25)$$

$$\int_{l_p}^{l_q} y^2 dl = \rho^3 \left[\left(\frac{y_c}{\rho} \right)^2 + \frac{1}{2} \right] (\phi_q - \phi_p) - 2 \frac{y_c}{\rho} (\cos \phi_q - \cos \phi_p) - \frac{1}{4} (\sin 2\phi_q - \sin 2\phi_p) \quad (26)$$

$$\int_{l_p}^{l_q} xy dl = \rho^3 \left[\frac{x_c y_c}{\rho^2} (\phi_q - \phi_p) - \frac{x_c}{\rho} (\cos \phi_q - \cos \phi_p) + \frac{y_c}{\rho} (\sin \phi_q - \sin \phi_p) - \frac{1}{4} (\cos 2\phi_q - \cos 2\phi_p) \right] \quad (27)$$

The equations given are all in a convenient form for easy numerical evaluation of the integrals.

It is not possible to give general formulae for dealing with cantilever whiskers, in which there may not be a simple relation between I and l . Whiskers of this type should be dealt with by calculating the section moment of inertia, I , and the co-ordinates (x, y) at a sufficient number of points to enable a graph of the functions to be plotted. A planimeter can then be used to evaluate the integrals by measuring the areas under the curves.

(7) PRACTICAL APPLICATIONS OF DESIGN THEORY

The transverse component of contact pressure expressed as a percentage of the normal component of pressure, has been calculated for the typical shapes whose dimensions are given in Figs. 4(a), 4(b) and 4(c), and the results are given below.

Single-crimp [Fig. 4(a)]	22.0%
Symmetric double-crimp [Fig. 4(b)]	7.6%
Corrected double-crimp [Fig. 4(c)]	2.6%

The dimensions of the corrected whisker shown in Fig. 4(c) were arrived at by the approximation $x_1y_1 = x_2y_2$, which is far from being an exact solution of $\int_0^L xydl = 0$ (x_1, y_1 and x_2, y_2 are the co-ordinates of the corners of the two crimps). This whisker is actually slightly overcorrected. If necessary, an exact

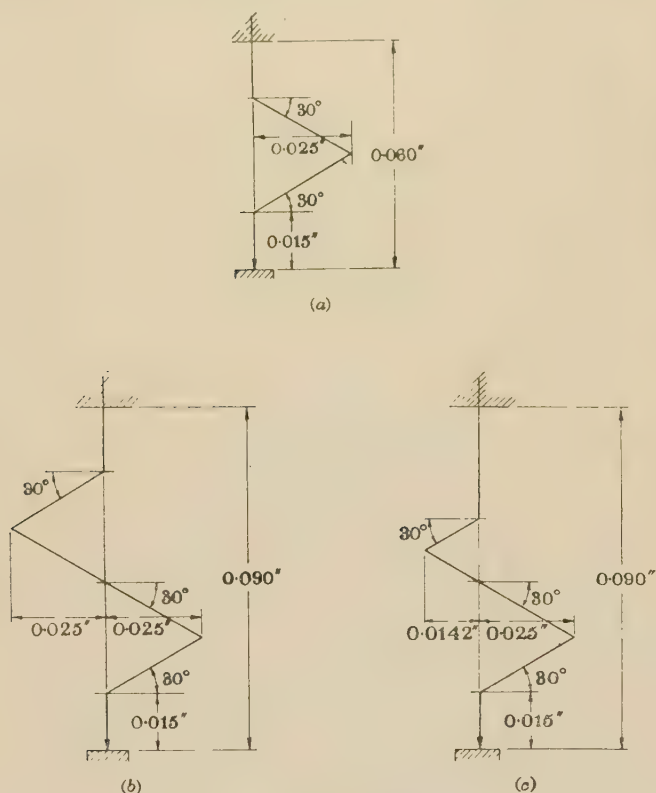


Fig. 4.—Dimensions of whiskers.

- (a) Single crimp.
(b) Symmetrical double crimp.
(c) Corrected double crimp.

solution could be arrived at by graphical interpolation of several trial calculations.

To summarize, the calculations show that single-crimp whiskers will produce a high transverse component of contact pressure, which can be reduced by using a double-crimp whisker of similar dimensions. Transverse pressure can be reduced still further (to zero if necessary) by a suitable reduction of the size of the crimp furthest from the crystal.

(8) EXPERIMENTAL RESULTS

Delicate measurements on actual-size whiskers were avoided by constructing the large-scale whisker model shown in Fig. 5. The model, consisting of $\frac{1}{8}$ in square-section steel bars connected by clamps designed to allow length-wise adjustment without loss of angular rigidity, was erected vertically to a height of approxi-

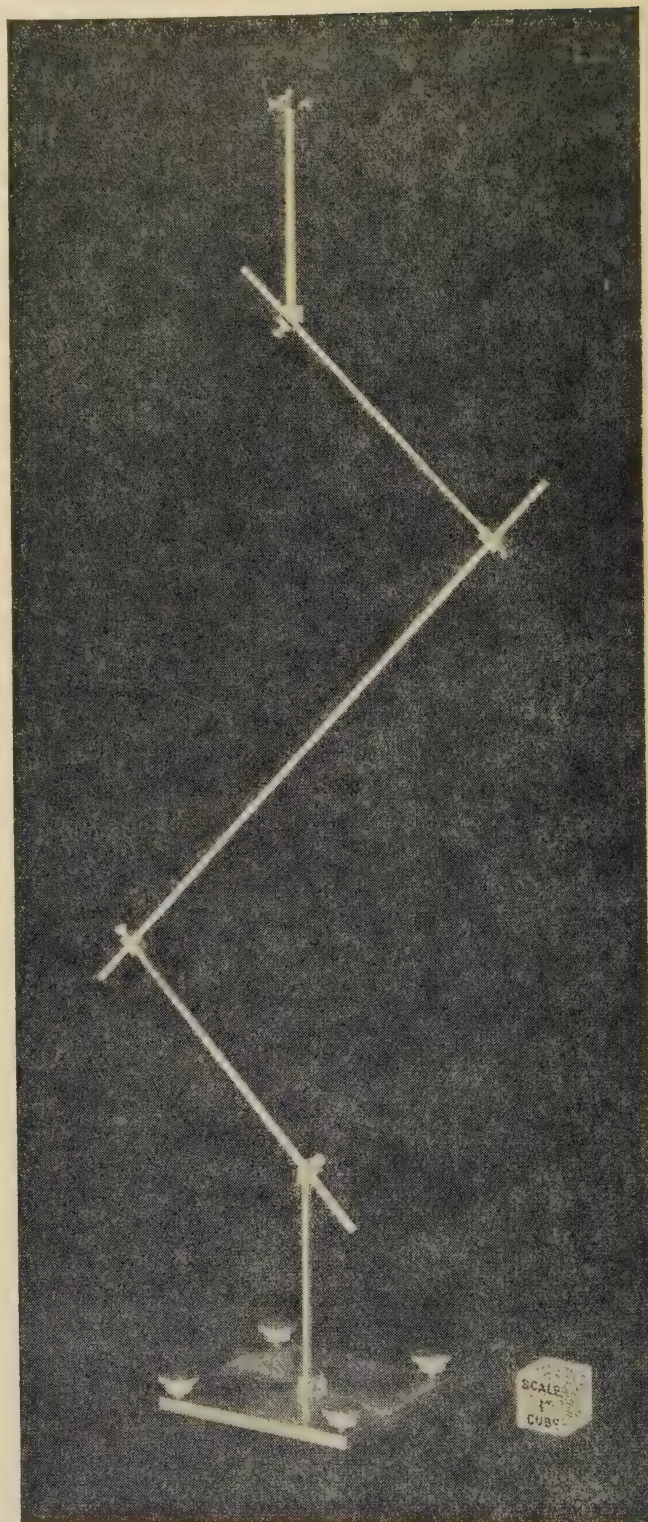


Fig. 5.—Large-scale whisker model.

mately 24 in on a rigid base, which represented the mounting point of the whisker. The top end of the uppermost limb represented the point of contact with the crystal.

Contact pressure normal to the face of the crystal was simulated by placing a weight on a collar at the top of the model. The transverse component of contact pressure was simulated by the

tension in a thread attached to the top of the model and pulling horizontally to one side. Suitable arrangements were made for measuring this tension. Vertical and horizontal deflections at the top of the model were observed with a measuring microscope.

Single-crimp, double-crimp, and corrected double-crimp arrangements were tested. Deflections were measured at zero transverse load, to test the deflection formulae (10)–(12), and also formula (14) in the special case of the corrected whisker. The tension in the thread required to restore the transverse deflection to zero was measured, to test eqns. (8) and (13). The vertical deflection under these conditions was measured to test eqn. (9). In this way all aspects of the theory were thoroughly tested.

These experiments confirmed theoretical predictions of the general behaviour of crimped whiskers, namely that double-crimp whiskers produce less transverse pressure than single-crimp whiskers, and that a suitable reduction in the size of the second crimp reduces the transverse pressure component to zero. Discrepancies between theoretical and actual values ranged up to 20%. Statistical analysis of the experimental data showed that these discrepancies had a high degree of significance, and an examination of the experimental procedure showed that they could be accounted for to a large extent by defects in the experiments (e.g. the distributed weight of the whisker model which is not taken into account by the theory, was appreciable compared with the applied load). In present practical devices, the distributed weights are insignificant compared with the applied pressures. This does not discount the theoretical basis, but

indicates that the experimental work may not apply to practical conditions.

It was concluded that the experiments were in sufficiently close agreement with the theory to justify using the theory for design purposes. There is good reason to believe that the theory is valid for plane whiskers of any shape, either in the form of crimped wires or cantilevers.

(9) CONCLUSIONS

The principal conclusions can be summarized as follows:

(a) A general theory of contact pressures and deflections of plane whiskers has been developed, and formulae for calculating deflections have been derived [eqns. (8) and (9)].

(b) The theoretical conditions for zero transverse contact pressure and/or zero transverse deflection have been determined [eqns. (12)–(14)].

(c) A theoretical study of crimped whiskers has shown that double-crimp whiskers have generally less transverse contact pressure than single-crimp whiskers, and for zero transverse contact pressure and deflection a double-crimp whisker with unequal crimps is required.

(d) The general behaviour of crimped whiskers, as predicted by theory, has been confirmed by experiments with a large-scale whisker model.

(10) ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of colleagues in the B.T.H. Research Laboratory, in particular that of Mr. E. R. C. Novotny in designing the model and carrying out experiments.

THE CALCULATION OF CYCLIC RATING FACTORS FOR CABLES LAID DIRECT OR IN DUCTS

By H. GOLDENBERG, M.Sc.

(The paper was first received 4th May, and in revised form 28th June, 1956. It was published as an INSTITUTION MONOGRAPH in October, 1956.)

SUMMARY

A simple method is presented for the rapid calculation of cyclic rating factors for cables laid direct in the ground and cables in ducts. The method may be applied without difficulty even in the case of complicated load curves, but for single-peaked load curves the calculation is especially simple. Tables of relevant functions are included as are additional tables of parameters for cables conforming to B.S. 480: 1954. Examples are given illustrating the ease and speed of the proposed method, together with the variation of cyclic rating factors with soil thermal constants.

LIST OF SYMBOLS

- a_x = Symbol defined in Section 6, given numerically by Tables 1 and 2.
 d = Outside diameter of the cable.
 f = Group rating factor for cables laid direct in the ground.
 g = Soil thermal resistivity.
 k = Ratio of outer cable temperature rise to core temperature rise, under steady conditions, for an isolated cable.
 \bar{k} = Value of k in the presence of remaining cables.
 l = Depth of the axis of the cable below the ground surface.
 n = Number of cores per cable.
 p = Constant proportion of peak load persisting for the remainder of a peaked load cycle, as defined in detail in Section 6.
 t = Time, seconds.
 A = Symbol defined in Section 3.
 B = Symbol defined in Appendix 11.1.
 D = Thermal diffusivity of the soil.
 G' = Internal thermal resistance between each core and sheath of each cable.
 G_e = External thermal resistance of a single cable, or of one cable of a group.
 G'_e = External thermal resistance of a cable whose outer diameter coincides with the inner duct wall.
 \bar{G}_e = External thermal resistance of one cable of a group.
 G_f = Thermal resistance of filling material between each lead sheath and armouring for an S.L.-type cable.
 G_s = Thermal resistance of the protective covering over the lead sheath of each cable, i.e. bedding, serving, etc.
 G_A = Thermal resistance between the outer cable surface and the inner duct wall.
 P = Losses per unit length of cable.
 Q = Quantity defined in Section 3.

Q' = Symbol defined in Section 3.

S = Thermal capacity of the cable.

T = Time, hours.

T_1, T_2 = Times in hours, defined in Section 3.

Y_0, Y_1, \dots, Y_{23} = Magnitudes of the appropriate hourly rectangular pulses into which the square of the daily cyclic load is decomposed. Each of the Y 's is expressed as a fraction of their maximum value.

$\alpha(T)$ = Ratio of transient outer cable temperature rise to steady-state outer cable temperature rise.

$\beta(T)$ = Symbol defined in Appendix 11.1.

$\theta_c(T)$ = Transient core temperature rise of a cable at a sufficiently large time T hours after the application of a step function load for the effect of the thermal capacity of the cable to be negligible.

λ = Sheath loss factor.

λ_1 = Armour loss factor.

(1) INTRODUCTION

Published tables of current ratings (maximum permissible currents) for British Standard cables are based on the continued application of a constant alternating or direct current. In practice the maximum load is usually applied only for a limited time, and the maximum current may then be increased without the maximum permitted conductor temperature rise being exceeded. Such increases in current rating, where the applied load is cyclic, are given by cyclic rating factors. The range of cyclic rating factors available until recently has been rather limited,^{1,2,3} although relevant formulae have been well known,^{2,3,4,5} as has been the use of an analogue computer for making the necessary calculations.⁶

Four recent papers^{7,8,9,10} describe the contribution of a working group of the American Institute of Electrical Engineers, on the subject of cyclic loading of pipe-type cables and buried lead-sheathed cables. The first of these covers ground mainly common to the remaining three, whilst each of the latter recommends a different procedure, the difference between the final results being negligible in view of the uncertainties regarding the thermal constants involved.¹¹ Only rectangular and sinusoidal loss cycles are covered^{8,9} by methods appropriate for routine application, complicated loss cycles being treated by tedious superposition methods.^{8,10,5}

For sinusoidal loss cycles the American methods provide a real advance in the rapid calculation of cyclic rating factors. The scope of immediate application is limited to load cycles of minimum loss factor 0.3. This is presumably because the effect of the thermal capacity of the cable itself is omitted from the calculation, and this effect became important at loss factors lower than 0.3, especially for large-diameter pipe-type cables. It is of interest to note that the experimental results used to check the calculations are those of an early E.R.A. report.¹²

The American work is put forward with application limited to cables laid direct, the cyclic ratings for ducted cables based on

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is based on Report Ref. F/T184 of the British Electrical and Allied Industries Research Association.

a large number of field measurements having been published earlier in that country.^{7, 13}

The present paper describes a new, simple and rapid method for the calculation of cyclic rating factors which may be applied without difficulty even in the case of very complex load curves. When the load curve assumes the form of a constant current for one portion of the cycle and a different constant current for the remainder of the cycle, the calculation becomes especially simple. Examples are given of the application of the method to cables laid direct in the ground and to cables in ducts.

In the body of the paper, the effect of the thermal capacity of the cable itself is assumed negligible as is usually the case. In Appendix 11.1 it is shown how this factor may be taken into account if desired, with little consequent increase in the amount of calculation.

(2) BASIC FORMULAE FOR THE TRANSIENT HEATING OF A BURIED CABLE

It has been shown by Whitehead and Hutchings^{2, 3} that the transient core-temperature rise $\theta_c(T)$ of a cable, at a sufficiently long time T hours after the application of a step function load for the effect of the thermal capacity of the cable to be negligible, is given by

$$\frac{\theta_c(T)}{\theta_c(\infty)} = 1 - k + k\alpha(T) \quad . \quad . \quad . \quad (1)$$

k = Ratio of outer cable temperature rise to core temperature rise, under steady conditions.

$\alpha(T)$ = Ratio of transient outer cable temperature rise to steady-state outer cable temperature rise.

The effect of the thermal capacity of most British Standard cables¹⁴ on consequent cyclic rating factors is negligible for times T of one hour or longer. Further, by neglecting the thermal capacity of the cable the rate of rise of the core temperature is somewhat overestimated and the corresponding cyclic rating factors are slightly underestimated, or on the safe side. Therefore, for the purposes of the present paper, consideration of the effect of the thermal capacity of a cable on the cyclic rating factor is relegated to Appendix 11.1.

Several formulae for $\alpha(T)$ have been presented, each based on different idealizations of the heat-flow problem. Following the usual practice we limit ourselves to those formulae treating the surface of the ground as an isothermal.

When the cable is considered as replaced by a continuous line source of heat at the cable centre, with a constant rate of heat liberation per unit time per unit length of the source, and when the cable is also assumed to possess the same thermal properties as the soil in which it is buried, the resultant formula for $\alpha(t)$ is readily obtainable from expressions given by Buller⁴ or Carslaw and Jaeger,^{15(a)} and is

$$\alpha(t) = \frac{-Ei\left(-\frac{d^2}{16Dt}\right) + Ei\left(-\frac{l^2}{Dt}\right)}{\log_e(16l^2/d^2)} \quad . \quad . \quad . \quad (2)$$

t = Time, sec.

d = Outside diameter of cable, cm.

l = Depth of the axis of the cable below the ground surface, cm.

D = Thermal diffusivity* of the soil, cm²/csec.

$-Ei(-x)$ is the exponential integral function defined by

$$-Ei(-x) = \int_x^\infty \frac{e^{-u}}{u} du$$

* Thermal diffusivity = $1/(\text{thermal resistivity} \times \text{density} \times \text{specific heat})$.

The assumption that the cable material possesses the thermal properties of the soil is not so unjustified in practice for the times under consideration as would appear at first sight.^{4, 16} Excellent tables of the exponential integral function have been published by the National Bureau of Standards.¹⁷ The formula (2) is often referred to as the exponential integral formula.

When the cable is considered as a continuous cylindrical surface source with a constant heat flux per unit area per unit time crossing its surface, an exact expression for $\alpha(T)$ may readily be obtained from Carslaw and Jaeger^{15(b)} or from formulae quoted by Buller.⁴ However, this exact expression for $\alpha(T)$ involves an infinite integral of a combination of Bessel functions, and is not suitable for extensive calculations. Buller⁴ has given a correction factor to be applied to the exponential integral formula (2) so as to give closer agreement with the formulae of Carslaw and Jaeger, but for small cable thermal capacity this correction factor is negligible. Whitehead and Hutchings^{2, 3} obtain an expression for $\alpha(T)$ for a continuous cylindrical surface source by integrating the effects of identical continuous line sources around a circle coinciding with the cable diameter. Their expression is obtained after approximations dependent on the depth of laying of the cable being much greater than the cable radius, and is

$$\alpha(t) = \frac{-Ei\left(-\frac{d^2}{8Dt}\right) + (\log_e 2) \exp\left(-\frac{d^2}{8Dt}\right) + Ei\left(-\frac{l^2}{Dt}\right)}{\log_e(16l^2/d^2)} \quad . \quad . \quad . \quad (3)$$

This formula has been used as a basis for the present work, but for the range of values of the parameters involved, the percentage difference between the values of $\alpha(t)$ given by the eqns. (2) and (3) is quite small, even for the most unfavourable cases. A more extended comparison between eqns. (2) and (3) is given in Appendix 11.2.

(3) BASIS OF METHOD OF CALCULATING CYCLIC RATING FACTORS

The heat dissipated in a cable may, with sufficient accuracy for the present purposes, be taken as directly proportional to the square of the applied current.

The core temperature rise of a buried cable at time T_2 hours after the application to the cable of a rectangular current pulse of duration $T_1 - T_2$ hours is $\theta_c(T_1) - \theta_c(T_2)$, and from eqn. (1) is given as a fraction of the steady-state core temperature rise for the same current by

$$\frac{\theta_c(T_1) - \theta_c(T_2)}{\theta_c(\infty)} = \left. \begin{aligned} &= [1 - k + k\alpha(T_1)] - [1 - k + k\alpha(T_2)] \\ &= k[\alpha(T_1) - \alpha(T_2)] \end{aligned} \right\} \text{ for } T_2 \geq 1$$

$$= [1 - k + k\alpha(T_1)] \quad \text{for } T_2 = 0$$

Hence the transient core temperature rise at time T_2 hours after the end of the last of a large number m of identical rectangular current pulses, each of duration $T_1 - T_2$, and repeated at intervals of 24 hours, has an upper limit given by

$$\sum_{m=0}^{\infty} \frac{[\theta_c(T_1 + 24m) - \theta_c(T_2 + 24m)]}{\theta_c(\infty)}$$

$$= k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)] \text{ for } T_2 \geq 1$$

$$= 1 - k + k\alpha(T_1) + k \sum_{m=1}^{\infty} [\alpha(T_1 + 24m) - \alpha(24m)] \text{ for } T_2 = 0$$

$$= 1 - k + k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(24m)] \text{ for } T_2 = 0 \quad . \quad (4)$$

Table 1

VALUES OF Q FOR SOIL THERMAL DIFFUSIVITY $0.02 \text{ cm}^2/\text{SEC}$

T_2	Cable diameter or inner duct diameter (in)																			T_2
	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	5.00	6.00	
0	0.539	0.504	0.475	0.450	0.429	0.410	0.392	0.374	0.358	0.344	0.330	0.316	0.303	0.291	0.279	0.268	0.257	0.237	0.200	0
1	0.071	0.076	0.080	0.084	0.087	0.090	0.093	0.095	0.097	0.099	0.101	0.103	0.105	0.107	0.108	0.109	0.110	0.112	0.115	1
2	0.045	0.048	0.051	0.054	0.056	0.058	0.060	0.061	0.062	0.064	0.065	0.067	0.068	0.069	0.070	0.071	0.072	0.074	0.077	2
3	0.034	0.037	0.039	0.041	0.043	0.044	0.046	0.047	0.048	0.049	0.050	0.051	0.052	0.053	0.054	0.055	0.055	0.057	0.060	3
4	0.029	0.031	0.033	0.034	0.035	0.037	0.038	0.039	0.040	0.041	0.042	0.043	0.043	0.044	0.045	0.046	0.046	0.047	0.050	4
5	0.025	0.027	0.028	0.029	0.031	0.032	0.033	0.033	0.034	0.035	0.036	0.037	0.038	0.039	0.039	0.040	0.040	0.042	0.044	5
6	0.022	0.024	0.025	0.026	0.028	0.028	0.029	0.030	0.031	0.032	0.032	0.033	0.034	0.034	0.035	0.035	0.036	0.037	0.039	6
7	0.020	0.022	0.023	0.024	0.025	0.026	0.027	0.027	0.028	0.029	0.030	0.030	0.031	0.031	0.032	0.032	0.033	0.034	0.036	7
8	0.019	0.020	0.021	0.022	0.023	0.024	0.025	0.025	0.026	0.027	0.027	0.028	0.028	0.029	0.029	0.030	0.030	0.031	0.033	8
9	0.017	0.019	0.020	0.021	0.022	0.022	0.023	0.024	0.024	0.025	0.025	0.026	0.026	0.027	0.028	0.028	0.029	0.029	0.031	9
10	0.016	0.018	0.019	0.020	0.020	0.021	0.022	0.022	0.023	0.024	0.024	0.024	0.025	0.025	0.026	0.026	0.027	0.027	0.029	10
11	0.015	0.017	0.018	0.018	0.019	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.024	0.024	0.025	0.025	0.025	0.026	0.027	11
12	0.015	0.016	0.017	0.018	0.018	0.019	0.019	0.020	0.021	0.021	0.022	0.022	0.022	0.023	0.023	0.024	0.024	0.025	0.026	12
13	0.014	0.015	0.016	0.017	0.017	0.018	0.019	0.019	0.020	0.020	0.021	0.021	0.022	0.022	0.022	0.023	0.023	0.024	0.025	13
14	0.014	0.015	0.016	0.016	0.017	0.017	0.018	0.018	0.019	0.019	0.020	0.020	0.021	0.021	0.021	0.022	0.022	0.023	0.024	14
15	0.013	0.014	0.015	0.016	0.016	0.017	0.017	0.018	0.018	0.019	0.019	0.019	0.020	0.020	0.021	0.021	0.021	0.022	0.023	15
16	0.013	0.014	0.015	0.015	0.016	0.016	0.017	0.017	0.018	0.018	0.018	0.019	0.019	0.020	0.020	0.020	0.021	0.021	0.022	16
17	0.012	0.013	0.014	0.015	0.015	0.016	0.016	0.017	0.017	0.017	0.018	0.018	0.018	0.019	0.019	0.019	0.020	0.020	0.022	17
18	0.012	0.013	0.014	0.014	0.015	0.015	0.016	0.016	0.017	0.017	0.017	0.018	0.018	0.018	0.018	0.019	0.019	0.020	0.021	18
19	0.012	0.012	0.013	0.014	0.014	0.015	0.015	0.016	0.016	0.017	0.017	0.017	0.017	0.018	0.018	0.018	0.019	0.019	0.020	19
20	0.011	0.012	0.013	0.013	0.014	0.014	0.015	0.015	0.016	0.016	0.016	0.017	0.017	0.017	0.018	0.018	0.018	0.019	0.020	20
21	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.015	0.015	0.015	0.016	0.016	0.016	0.017	0.017	0.017	0.018	0.018	0.019	21
22	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.015	0.015	0.015	0.016	0.016	0.016	0.017	0.017	0.017	0.017	0.018	0.019	22
23	0.010	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.015	0.015	0.015	0.016	0.016	0.016	0.016	0.017	0.017	0.017	0.018	23

Table 2

VALUES OF Q FOR SOIL THERMAL DIFFUSIVITY $0.01 \text{ cm}^2/\text{SEC}$

T_2	Cable diameter or inner duct diameter (in)																			T_2
	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	5.00	6.00	
0	0.480	0.441	0.409	0.381	0.357	0.335	0.316	0.297	0.279	0.263	0.248	0.234	0.220	0.208	0.196	0.185	0.174	0.153	0.119	0
1	0.073	0.079	0.083	0.087	0.090	0.093	0.096	0.098	0.100	0.102	0.103	0.104	0.105	0.106	0.107	0.107	0.108	0.108	0.107	1
2	0.048	0.051	0.054	0.057	0.059	0.061	0.063	0.064	0.066	0.067	0.068	0.070	0.071	0.072	0.073	0.073	0.074	0.076	0.077	2
3	0.037	0.040	0.042	0.044	0.046	0.048	0.049	0.050	0.052	0.053	0.054	0.055	0.056	0.057	0.057	0.058	0.059	0.060	0.063	3
4	0.031	0.034	0.036	0.037	0.039	0.040	0.041	0.042	0.043	0.044	0.045	0.046	0.047	0.048	0.048	0.049	0.050	0.051	0.054	4
5	0.027	0.029	0.031	0.033	0.034	0.035	0.036	0.037	0.038	0.039	0.040	0.041	0.041	0.042	0.043	0.044	0.044	0.045	0.048	5
6	0.025	0.027	0.028	0.029	0.030	0.032	0.033	0.034	0.034	0.035	0.036	0.037	0.038	0.038	0.039	0.039	0.040	0.041	0.043	6
7	0.023	0.025	0.026	0.027	0.028	0.029	0.030	0.031	0.032	0.033	0.033	0.034	0.035	0.035	0.036	0.036	0.037	0.038	0.040	7
8	0.021	0.023	0.024	0.025	0.026	0.027	0.028	0.029	0.030	0.030	0.031	0.032	0.032	0.033	0.033	0.034	0.034	0.035	0.037	8
9	0.020	0.022	0.023	0.024	0.025	0.026	0.026	0.027	0.028	0.029	0.029	0.030	0.030	0.031	0.031	0.032	0.033	0.033	0.035	9
10	0.019	0.020	0.022	0.023	0.024	0.024	0.025	0.026	0.026	0.027	0.028	0.028	0.029	0.029	0.030	0.030	0.031	0.032	0.033	10
11	0.018	0.020	0.021	0.022	0.022	0.023	0.024	0.025	0.025	0.026	0.026	0.027	0.027	0.028	0.028	0.029	0.029	0.030	0.032	11
12	0.017	0.019	0.020	0.021	0.022	0.022	0.023	0.024	0.024	0.025	0.025	0.026	0.026	0.027	0.027	0.028	0.028	0.029	0.031	12
13	0.017	0.018	0.019	0.020	0.021	0.021	0.022	0.023	0.023	0.024	0.024	0.025	0.025	0.026	0.026	0.027	0.027	0.028	0.029	13
14	0.016	0.017	0.018	0.019	0.020	0.021	0.021	0.022	0.023	0.023	0.024	0.024	0.025	0.025	0.025	0.026	0.026	0.027	0.028	14
15	0.016	0.017	0.018	0.019	0.019	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.024	0.024	0.025	0.025	0.025	0.026	0.028	15
16	0.015	0.016	0.017	0.018	0.019	0.019	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.023	0.024	0.024	0.025	0.025	0.027	16
17	0.015	0.016	0.017	0.017	0.018	0.019	0.019	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.023	0.024	0.024	0.025	0.026	17
18	0.014	0.015	0.016	0.017	0.018	0.018	0.019	0.020	0.020	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.023	0.024	0.025	18
19	0.014	0.015	0.016	0.016	0.017	0.018	0.019	0.019	0.019	0.020	0.020	0.021	0.021	0.022	0.022	0.023	0.023	0.023	0.025	19
20	0.014	0.015	0.015	0.016	0.017	0.018	0.018	0.019	0.019	0.020	0.020	0.020	0.021	0.021	0.022	0.022	0.022	0.023	0.024	20
21	0.013	0.014	0.015	0.016	0.017	0.017	0.018	0.018	0.019	0.019	0.020	0.020	0.020	0.021	0.021	0.021	0.022	0.022	0.024	21
22	0.013	0.014	0.015	0.016	0.016	0.017	0.017	0.018	0.018	0.019	0.019	0.019	0.020	0.020	0.021	0.021	0.021	0.022	0.023	22
23	0.013	0.014	0.015	0.015	0.016	0.016	0.017	0.017	0.018	0.018	0.019	0.019	0.019	0.020	0.020	0.020	0.021	0.022	0.023	23

$\alpha(T)$ being given by the right-hand side of eqn. (3) with $t = 3600T$.

This infinite series is very slowly convergent, and does not appear to be directly summable in terms of analytic functions. However, it can be summed numerically by means of an artifice depending on the fact that if $T_1 - T_2 = 1$ hour, say, then 24 such sets of pulses, suitably displaced and each of the same amplitude, constitute a continuous load.

Consider the sum of the 24 values of

$$Q = \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)] \quad (5)$$

for $T_1 - T_2 = 1$ and $T_2 = 0(1)23$ for any given values of the parameters D , d and l . Each value of Q represents the outer cable temperature rise at a time T_2 hours after the repeated daily application of an hourly load pulse, expressed as a ratio of the ultimate outer cable temperature rise on application to the cable of a constant load of amplitude equal to that of each of the pulses. Hence the sum of these 24 values is unity. Suppose then that the 24 values of

$$Q' = \sum_{m=0}^n [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)]$$

for $T_1 - T_2 = 1$ and $T_2 = 0(1)23$ are calculated, corresponding to $(n + 1)$ successive days with applied load. The sum of these 24 values will not be unity, but will be close to unity provided that n is chosen to be sufficiently large. The procedure adopted in the present paper has been to divide the difference between the sum of these 24 values of Q' and unity by 24, and to add this constant amount to each of the 24 values of Q' , so that the sum of the consequent values is unity. The values so obtained have been taken as representing the required values of Q , provided that there is no change in these values (to the required accuracy) when n is increased by unity before subdivision. This process of successive approximation has been found to be rapidly convergent, the calculation being necessary for a maximum of 5 days' load.

The core temperature rise at any time during the cycle for a cable carrying a daily cyclic load can then readily be calculated by the addition of the 24 terms, each term being weighted by the amplitude of the corresponding hourly rectangular pulse into which the square of the daily cyclic load is decomposed.

Denoting $Q_r = Q$ for $T_1 - T_2 = 1$ and $T_2 = r$ hours, and further denoting the magnitudes of the appropriate hourly rectangular pulses into which the square of the daily cyclic load is decomposed by Y_0, Y_1, \dots, Y_{23} , where each of the Y 's is expressed as a fraction of their maximum value, the cyclic load as a whole may therefore be multiplied by the following cyclic rating factor for the same core temperature rise:

Cyclic rating factor

$$= \frac{1}{\sqrt{[Y_0(1 - k + kQ_0) + Y_1kQ_1 + Y_2kQ_2 + \dots + Y_{23}kQ_{23}]}}$$

$$= \frac{1}{\sqrt{[(1 - k)Y_0 + k(Y_0Q_0 + Y_1Q_1 + \dots + Y_{23}Q_{23})]}}$$

Denoting

$$A = Y_0Q_0 + Y_1Q_1 + \dots + Y_{23}Q_{23} = \sum_{r=0}^{23} Y_r Q_r = \sum_{T_2=0}^{23} Y(T_2)Q(T_2)$$

$$\text{Cyclic rating factor} = \frac{1}{\sqrt{[(1 - k)Y_0 + kA]}} \quad (6)$$

The direct application of this formula is later illustrated by practical examples.

Calculated values of Q for $T_2 = 0(1)23$, and for cable diameters of 0.5 (0.25) 4.5, 5.0, 6.0 in and depth of laying 3 ft are given in Table 1 for soil thermal diffusivity of $0.02 \text{ cm}^2/\text{sec}$ and in Table 2 for soil thermal diffusivity of $0.01 \text{ cm}^2/\text{sec}$.

The range of cable diameters has been extended to 6 in to permit utilization for cables in ducts. Two values of thermal diffusivity are covered, because although for the calculation of cyclic rating factors $0.02 \text{ cm}^2/\text{sec}$ has been used as a safe upper value for the majority of British soils,^{2,3} $0.01 \text{ cm}^2/\text{sec}$ may in many cases be a more realistic upper limit giving a higher cyclic rating factor in a soil of given thermal resistivity.

Safe cyclic rating factors for cables in ducts will be obtained on the assumption that, for the purposes of eqn. (1), the inner wall of the duct is taken as the outer diameter of an equivalent cable, the duct material having a thermal resistivity not materially different from that of the majority of British soils.

Correction factors to be applied to Tables 1 and 2 to take account of cable depths of burial other than 3 ft do not appear to be readily obtainable. However, one important property of eqn. (3) may be utilized directly, namely that the substitution $l = cl$, $d = cd$ and $D = c^2D$ leaves the equation invariant, and hence also leaves Tables 1 and 2 invariant. Therefore the values corresponding to Table 1 for a soil thermal diffusivity D of $0.02 \text{ cm}^2/\text{sec}$ but for a cable depth of burial $\sqrt{2} \times 3 = 4.24 \text{ ft}$ may readily be obtained from Table 2, which has been calculated for $D = 0.01$ and a 3 ft depth of burial, provided that the correct cable diameter is divided by a factor of $\sqrt{2}$ before entry into Table 2. Similarly, the values corresponding to Table 2 for a soil thermal diffusivity D of $0.01 \text{ cm}^2/\text{sec}$ but for a cable depth of burial $3/\sqrt{2} = 2.12 \text{ ft}$ may readily be obtained from Table 1, which has been calculated for $D = 0.02$ and a 3 ft depth of burial, provided that the correct cable diameter is multiplied by a factor of $\sqrt{2}$ before entry into Table 1. This point will be illustrated in a later example.

Table 3

FIGURES TO UTILIZE A MORE ACCURATE REPRESENTATION OF THE APPLIED LOAD IN THE NEIGHBOURHOOD OF ITS PEAK. (SEE APPENDIX 11.1)

Cable diameter	Two rows to replace row for $T_2 = 0$ in Table 1		Two rows to replace row for $T_2 = 0$ in Table 2	
	$T_2 = 0$	$T_2 = 0.5$	$T_2 = 0$	$T_2 = 0.5$
in				
0.50	0.473	0.066	0.413	0.067
0.75	0.433	0.071	0.369	0.072
1.00	0.400	0.075	0.333	0.076
1.25	0.372	0.078	0.302	0.079
1.50	0.348	0.081	0.275	0.082
1.75	0.326	0.084	0.251	0.084
2.00	0.306	0.086	0.230	0.086
2.25	0.287	0.087	0.210	0.087
2.50	0.269	0.089	0.191	0.088
2.75	0.253	0.091	0.174	0.089
3.00	0.238	0.092	0.159	0.089
3.25	0.223	0.093	0.144	0.090
3.50	0.209	0.094	0.131	0.089
3.75	0.196	0.095	0.119	0.089
4.00	0.184	0.095	0.108	0.088
4.25	0.172	0.096	0.098	0.087
4.50	0.161	0.096	0.088	0.086
5.00	0.141	0.096	0.071	0.082
6.00	0.106	0.094	0.046	0.073

Tables 1 and 2 cover only intervals of 1 h. Cases may arise in which the square of the current varies appreciably during the hour prior to the expected time of maximum core temperature. In such cases it may be advantageous to consider times $T_2 = 0, \frac{1}{2}, 1, 2, 3 \dots 23$ hours. Table 3 contains the data required to extend Tables 1, 2 to cover this additional subdivision. When Table 3 is utilized it may also be desirable to consider the effect of the thermal capacity of the cable (see Appendix 11.1).

(4) THE FACTOR k AND THE EFFECT OF ENVIRONMENT AND GROUPING OF CABLES

The evaluation of a cyclic rating factor for any particular cable and applied cyclic load requires the evaluation of k . Neglecting

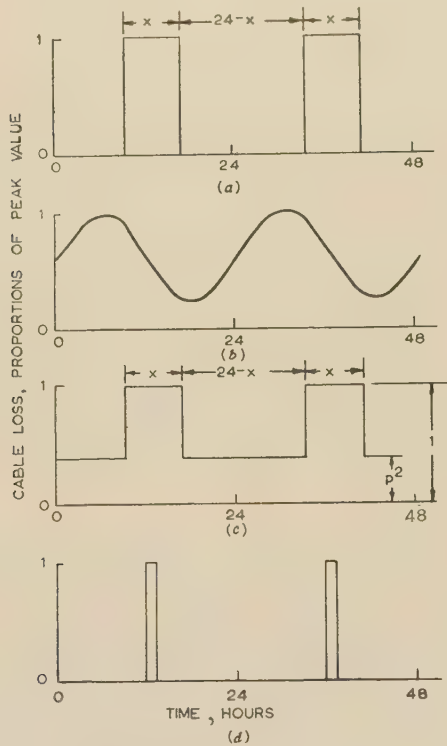


Fig. 1.—Basic loss cycles.

- (a) Rectangular cycle.
- (b) Sinusoidal cycle.
- (c) Peaked cycle.
- (d) Hour pulse cycle.

dielectric loss it may be shown that for a single cable, or group of cables, laid direct in the ground, k is given by

$$k = \frac{G_e}{G_e + G_s + \left[\left(\frac{G'}{n} - \lambda_1 \frac{G_s}{2} \right) / (1 + \lambda + \lambda_1) \right]} \quad (7)$$

while for cables laid in ducts, k is given by

$$k = \frac{G'_e}{G'_e + G_A + G_s + \left[\left(\frac{G'}{n} - \lambda_1 \frac{G_s}{2} \right) / (1 + \lambda + \lambda_1) \right]} \quad (8)$$

n = Number of cores per cable.

G' = Internal thermal resistance between each core and the sheath of each cable, thermal ohm-cm.

G = Thermal resistance of the protective covering (i.e. bedding, serving, etc.) over the lead sheath of each cable.

G_e = External thermal resistance of each cable of the group, thermal ohm-cm.

G'_e = External thermal resistance of a cable whose outer diameter coincides with the inner duct wall, thermal ohm-cm.

G_A = Thermal resistance between the outer cable surface and the inner duct wall, thermal ohm-cm.

λ = Sheath loss factor.

λ_1 = Armour loss factor.

For S.L. cables, eqns. (7) and (8) are valid in a slightly modified form. The factor λ_1 should be replaced by $\lambda_1(1 - \lambda)$, while G' is replaced by $G' + G_f$, where G_f is the thermal resistance of

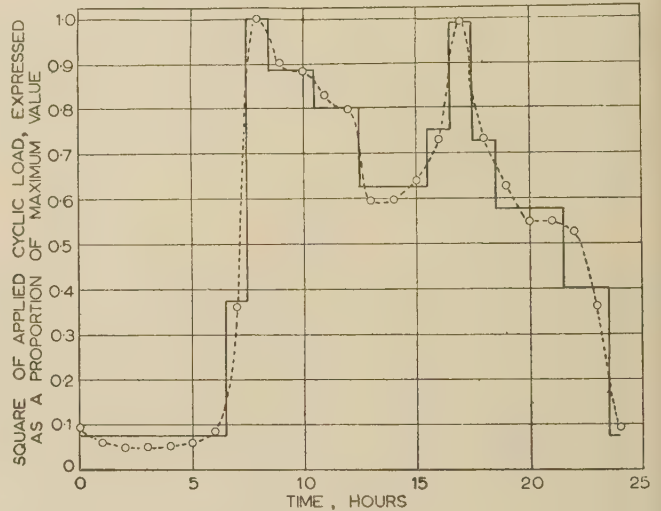


Fig. 2.—Square of applied cyclic load and equivalent step-function cyclic load.

- Square of applied cyclic load.
- Equivalent step function cyclic load.

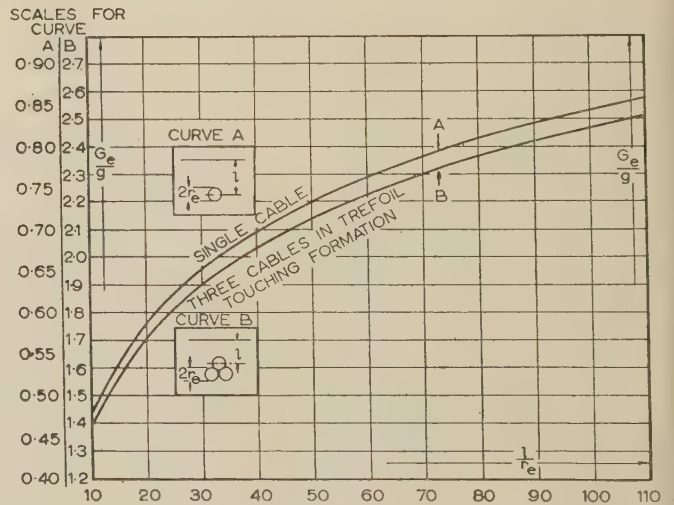


Fig. 3.—Graphs to determine the external thermal resistance of a single cable or of one of three cables in trefoil touching formation (cables laid direct in the ground).

g = Soil thermal resistivity.

l = Depth of (upper) cable axis below ground surface.

r_e = Cable radius.

Curve A.

G_e = External thermal resistance of single cable laid direct in ground.

Curve B.

G_B = External thermal resistance of one of three cables in trefoil touching formation, vertex upwards, laid direct in ground.

filling material between each lead sheath and armouring. Further G_e assumes the value of the thermal resistance of the protective covering over the armour of a cable.

Formulae for the evaluation of the above thermal resistances and loss factors for any cable and any cable configuration have been given by Whitehead and Hutchings.^{2,3} For convenience of calculation a non-dimensional graph is presented in Fig. 3 enabling G_e to be readily evaluated for a single multi-core cable or for three single-core cables in trefoil touching formation.

The value of k may also be found from the formula

$$k = \frac{PG_e}{\theta_c}$$

where θ_c is the core temperature rise for the cable working at a current giving losses of P watts per centimetre length of cable, and G_e is the cable external thermal resistance per centimetre. This formula is convenient for use with cables for voltages above 33 kV, the dimensions of which have not yet been standardized,

Table 4

CALCULATED VALUES OF k FOR REPRESENTATIVE AREAS OF CABLES CONFORMING TO B.S. 480: 1954.
CABLES LAID DIRECT IN THE GROUND

(Soil thermal resistivity, 120° C-cm/watt)

Nominal conductor area	1.1 kV†	3.3 kV	6.6 kV	11 kV	22 kV
in ²					
*0.007	0.48	—	—	—	—
*0.0225	0.54	0.57	0.55	0.54	—
0.0225	0.57	0.60	0.58	0.57	—
0.04	0.62	0.63	0.61	0.60	0.54
0.10	0.67	0.69	0.67	0.65	0.59
0.30	0.75	0.76	0.74	0.72	0.68
0.50	0.77	0.79	0.78	0.76	—
0.75	0.79	—	—	—	—
1.00	0.81	—	—	—	—

(a)

Nominal conductor area	1.1 kV†	3.3 kV	6.6 kV	11 kV	22 kV	33 kV
in ²						
*0.007	0.71	—	—	—	—	—
*0.0225	0.77	0.78	0.75	0.73	—	—
0.10	0.84	0.85	0.83	0.81	0.77	0.73
*0.25	0.88	0.88	0.86	0.85	0.81	0.79
*0.50	0.90	0.91	0.89	0.88	0.85	0.83
*0.75	0.90	0.92	0.91	0.90	0.87	0.85
*1.00	0.91	0.93	0.92	0.91	0.88	0.87

(e)

Nominal conductor area	1.1 kV†	3.3 kV	6.6 kV	11 kV	22 kV
in ²					
*0.007	0.53	—	—	—	—
*0.0225	0.59	0.62	0.60	0.58	—
0.0225	0.64	0.65	0.63	0.61	—
0.04	0.67	0.69	0.66	0.64	0.57
0.10	0.73	0.74	0.72	0.70	0.63
0.30	0.80	0.80	0.78	0.76	0.69
0.50	0.81	0.83	0.81	0.79	—
0.75	0.82	—	—	—	—
1.00	0.83	—	—	—	—

(b)

Nominal conductor area	11 kV	22 kV	33 kV
in ²			
0.0225	0.59	—	—
0.04	0.63	0.60	—
0.10	0.69	0.64	0.62
0.25	0.74	0.69	0.68
0.50	0.78	0.74	0.73

(f)

Nominal conductor area	22 kV	33 kV
in ²		
0.04	0.63	—
0.10	0.68	0.65
0.30	0.74	0.72
0.50	0.77	0.76

(c)

Nominal conductor area	11 kV	22 kV	33 kV
in ²			
0.0225	0.63	—	—
0.04	0.67	0.63	—
0.10	0.73	0.67	0.64
0.25	0.78	0.72	0.70
0.50	0.81	0.75	0.74

(g)

Nominal conductor area	Armoured	Unarmoured
in ²		
*0.007	0.48	0.53
0.0225	0.54	0.60
0.04	0.58	0.64
0.10	0.64	0.70
0.30	0.71	0.77
0.50	0.74	0.78
0.75	0.76	0.79
1.00	0.78	0.80

(d)

- (a) Three-core belted armoured and served cables.
(b) Three-core belted unarmoured and served cables.
(c) Three-core S.L. armoured and served cables.
(d) Twin belted and served cables 1.1 kV.
(e) Three single-core served cables in trefoil touching formation.
(f) Three-core screened armoured and served cables.
(g) Three-core screened unarmoured and served cables.

* Circular conductors.

† Depth of cable surface below ground: 36 in for all cable voltages above 1.1 kV; 25.5 in for cable voltage = 1.1 kV.

All cables have added sheath protection as specified in B.S. 480: 1954.

but for which the values P and θ_c are usually accurately known, while G_c may be obtained from the external dimensions and depth of cable in a given soil, Fig. 3 being utilized where applicable.

Values of k calculated for representative areas for all voltages for cables conforming with the latest British Standard¹⁴ are given in detail in Tables 4 and 5 for cables laid direct and cables in ducts, respectively. These values may be used directly where applicable. They are based on a soil thermal resistivity of 120° C-cm/watt, a protective-covering thermal resistivity of 500° C-cm/watt and a dielectric thermal resistivity of 750° C-cm/watt for 1.1 kV cables

and 550° C-cm/watt for higher-voltage cables up to and including 33 kV. These calculated values of k may readily be adjusted to correspond to any different soil thermal resistivity g ° C-cm/watt, by the use of the formula

$$k(g) = \frac{1}{1 + \frac{120 [1 - k(120)]}{g k(120)}} \quad (9)$$

Eqn. (9) may also be taken as sufficiently accurate to be applicable to cables in ducts, although G_A is not strictly a constant.

Table 5

CALCULATED VALUES OF k FOR REPRESENTATIVE AREAS OF CABLES CONFORMING TO B.S. 480: 1954. CABLES LAID IN DUCTS
(Soil thermal resistivity, 120° C-cm/watt)

Nominal conductor area	1.1 kV	3.3 kV	6.6 kV	11 kV	22 kV
in ²					
*0.007	0.27	—	—	—	—
*0.0225	0.31	0.33	0.33	0.32	—
0.0225	0.32	0.33	0.34	0.33	—
0.04	0.35	0.36	0.36	0.35	0.34
0.10	0.40	0.41	0.41	0.40	0.39
0.30	0.47	0.48	0.47	0.45	0.45
0.50	0.50	0.51	0.50	0.49	—
0.75	0.54	—	—	—	—
1.00	0.56	—	—	—	—

(a)

Nominal conductor area	1.1 kV	3.3 kV	6.6 kV	11 kV	22 kV
in ²					
*0.007	0.26	—	—	—	—
*0.0225	0.30	0.32	0.32	0.31	—
0.0225	0.31	0.32	0.32	0.31	—
0.04	0.33	0.34	0.35	0.34	0.34
0.10	0.38	0.39	0.39	0.38	0.38
0.30	0.46	0.46	0.46	0.45	0.43
0.50	0.49	0.50	0.50	0.46	—
0.75	0.51	—	—	—	—
1.00	0.54	—	—	—	—

(b)

Nominal conductor area	22 kV	33 kV
in ²		
0.04	0.40	—
0.10	0.44	0.46
0.30	0.50	0.52
0.50	0.54	0.55

(c)

Nominal conductor area	Armoured	Unarmoured
in ²		
*0.007	0.27	0.26
0.0225	0.30	0.29
0.04	0.33	0.31
0.10	0.37	0.36
0.30	0.44	0.42
0.50	0.48	0.46
0.75	0.50	0.49
1.00	0.53	0.51

(d)

Nominal conductor area	1.1 kV	3.3 kV	6.6 kV	11 kV	22 kV	33 kV
in ²						
*0.007	0.42	—	—	—	—	—
*0.0225	0.47	0.48	0.48	0.46	—	—
*0.10	0.56	0.56	0.56	0.54	0.53	0.54
*0.25	0.62	0.62	0.61	0.60	0.59	0.60
*0.50	0.66	0.66	0.66	0.65	0.64	0.65
*0.75	0.69	0.69	0.69	0.68	0.67	0.68
*1.00	0.71	0.72	0.72	0.70	0.70	0.71

(e)

Nominal conductor area	11 kV	22 kV	33 kV
in ²			
0.0225	0.35	—	—
0.04	0.37	0.39	—
0.10	0.42	0.41	0.42
0.25	0.47	0.46	0.48
0.50	0.51	0.50	0.51

(f)

Nominal conductor area	11 kV	22 kV	33 kV
in ²			
0.0225	0.34	—	—
0.04	0.36	0.37	—
0.10	0.40	0.41	0.41
0.25	0.45	0.45	0.46
0.50	0.49	0.49	0.51

(g)

- (a) Three-core belted armoured and served cables.
- (b) Three-core belted unarmoured and served cables.
- (c) Three-core S.L. armoured and served cables.
- (d) Twin belted and served cables 1.1 kV.
- (e) Three single-core served cables in trefoil touching formation.
- (f) Three-core screened armoured and served cables.
- (g) Three-core screened unarmoured and served cables.

* Circular conductors.

Depth of duct surface below ground: 30 in for all cable voltages up to 22 kV; 36 in for cable voltage of 33 kV.

All cables have added sheath protection as specified in B.S. 480: 1954.

Note.—Slight irregularities in above figures are largely due to the basis of duct choice of Ref. F/T183, namely:

4 in ducts for cables up to 2½ in diameter.

5 in ducts for cables above 2½ to 3½ in diameter.

6 in ducts for cables above 3½ in diameter.

For a single cable, or one of a sufficiently simple group of cables the dependence of k on the depth of cable burial may readily be ascertained from eqn. (7) or (8), as G_e or G'_e , the only terms substantially dependent on cable depth l , may be simply expressed in terms of that depth l .

There is also an interesting and useful connection between group rating factors and the appropriate value of k for such groups of cables. The group rating factor f for cables laid direct in the ground is given by

$$f^2 = \frac{\frac{G'}{n} + \left(G_e + \frac{G_s}{2}\right)(1 + \lambda + \lambda_1) + \frac{G_s}{2}(1 + \lambda)}{\frac{\bar{G}'}{n} + \left(\bar{G}_e + \frac{G_s}{2}\right)(1 + \lambda + \lambda_1) + \frac{G_s}{2}(1 + \lambda)} \quad (10)$$

where \bar{G}_e and G_e are the external thermal resistances of a single cable of a group in the presence and absence respectively of the remaining cables, while the appropriate value of k is given by

$$\bar{k} = \frac{\bar{G}_e}{\bar{G}_e + G_s + \left[\left(\frac{G'}{n} - \frac{\lambda_1 G_s}{2}\right)/(1 + \lambda + \lambda_1)\right]} \quad (7a)$$

for a single cable in the presence of remaining cables, and by eqn. (7) for a single cable in the absence of remaining cables.

From eqns. (10), (7a) and (7) we obtain the relation

$$\bar{k} = 1 - f^2(1 - k) \quad (11)$$

giving \bar{k} in terms of the group rating factor f and the value of k for a single cable in the absence of remaining cables. Hence, if the appropriate group rating factor is available,¹⁹ the calculation of \bar{G}_e may be avoided.

(5) EXAMPLES OF CYCLIC RATING FACTOR CALCULATIONS

The cyclic load chosen for illustration purposes is based on that given as the national load¹⁸ for 1st December, 1947. The load is assumed to be at constant voltage and power factor, so that variation of load represents variation of current. We are directly concerned only with the square of current expressed as a proportion of its peak value, and this ratio is plotted in Fig. 2 (dotted lines), together with an equivalent load curve of rectangular steps. Numerical values corresponding to these two curves are given in Table 6. It should be noticed that the two curves have approximately the same mean value, while the number of rectangular steps chosen has deliberately been reduced from a possible maximum of 24 to facilitate subsequent calculation. All equivalent curve peaks are considered as persisting for at least one hour. In cases when the time of maximum temperature is not obvious it may be necessary to perform the calculation more than once using each possible time of highest temperature as zero time, but in the present case this is obviously at 17.5 hours. In particular, in cases where the load falls slowly after a peak, the temperature may continue to rise for a short time after peak load.

(5.1) Example 1—Isolated Circuit Laid Direct, Standard Conditions of Laying, Various Soil Constants

Cable under consideration.—One isolated circuit comprising a single-core 1.00 in² 33 kV p.i.l.s. cables, served and with added sheath protection, conforming to B.S. 480: 1954.

Method of laying.—Direct in the ground in trefoil touching formation at depth 36 in of upper cable axis below ground surface.

VOL. 104, PART C.

Table 6

LOAD DATA CORRESPONDING TO NATIONAL LOAD CURVE OF FIG. 2

T	Cyclic load/peak load at time T	(Cyclic load/peak load) ² at time T
hours		
0	0.302	0.091
1	0.247	0.061
2	0.227	0.052
3	0.232	0.054
4	0.235	0.056
5	0.246	0.061
6	0.290	0.084
7	0.600	0.360
8	1.000	1.000
9	0.950	0.902
10	0.940	0.844
11	0.910	0.828
12	0.892	0.796
13	0.770	0.593
14	0.772	0.596
15	0.800	0.640
16	0.853	0.728
17	0.996	0.992
18	0.853	0.728
19	0.790	0.624
20	0.740	0.548
21	0.740	0.548
22	0.722	0.521
23	0.600	0.360

(a)

T	Magnitude of step function at time T
hours	
0–6.5	0.075
6.5–7.5	0.375
7.5–8.5	1.000
8.5–10.5	0.885
10.5–12.5	0.800
12.5–15.5	0.625
15.5–16.5	0.75
16.5–17.5	0.99
17.5–18.5	0.725
18.5–21.5	0.575
21.5–23.5	0.40
23.5–24	0.075

(b)

(a) National load curve.

(b) Step function equivalent to (cyclic load per peak load)².

Soil constants.—Soil thermal diffusivity $D = 0.02 \text{ cm}^2/\text{sec}$. Soil thermal resistivity $g = 120^\circ \text{ C-cm/watt}$.

Load.—As described above, illustrated in Fig. 2 and tabulated in Table 6.

Additional data required.—Overall diameter of each cable = 2.41 in (from B.S. 480: 1954). At above depth of burial and soil thermal resistivity, k is equal to 0.87 (from Table 4).

Table 7

CALCULATION OF CYCLIC RATING FACTOR FOR EXAMPLE 1.
NATIONAL LOAD CURVE OF FIG. 2 APPLIED TO ONE ISOLATED
CIRCUIT COMPRISING THREE SINGLE-CORE 1.00IN² 33kV
CABLES LAID DIRECT IN THE GROUND IN TREFOIL TOUCHING
FORMATION

(1)	(2)	(3)	(4)	(5)
Time*	Y step function equivalent to (cyclic load per peak load) ²	Q-factors for $D = 0.02$, $d = 2.50$ from Table 1	Summation of (3) over consecutive constant values of (2)	(2) × (4)
hours				
0-1	0.99	0.358	0.358	0.354
1-2	0.75	0.097	0.097	0.073
2-3	0.625	0.062	0.150	0.094
3-4	0.625	0.048		
4-5	0.625	0.040		
5-6	0.80	0.035	0.066	0.053
6-7	0.80	0.031		
7-8	0.885	0.028		
8-9	0.885	0.026	0.054	0.048
9-10	1.00	0.024		
10-11	0.375	0.023		
11-12	0.075	0.022	0.023	0.009
12-13	0.075	0.021		
13-14	0.075	0.020		
14-15	0.075	0.019	0.135	0.010
15-16	0.075	0.018		
16-17	0.075	0.018		
17-18	0.075	0.017	0.033	0.013
18-19	0.40	0.017		
19-20	0.40	0.016		
20-21	0.575	0.016	0.046	0.026
21-22	0.575	0.015		
22-23	0.575	0.015		
23-24	0.725	0.015	0.015	0.011
Total = A =				0.715

* Time measured prior to expected time of maximum core temperature.

Depth of upper cable axis below ground surface = 36in.

Outer diameter d of each cable of trefoil formation = 2.41 in.

Soil thermal diffusivity $D = 0.02 \text{ cm}^2/\text{sec}$.

Soil thermal resistivity $g = 120^\circ \text{C-cm/watt}$.

Value of k from Table 4, $k = 0.87$.

Value of column (2) for time 0-1 hour = $Y_0 = 0.99$.

Cyclic rating factor = $\frac{1}{\sqrt{[(1-k)Y_0 + kA]}} = 1.15$.

Method of Calculation (see Table 2).

Table 7 is a proposed standard table in which the time in hours has already been printed in column 1.

(i) Enter in column 2 the values of Y , namely the step function equivalent to (cyclic load per peak load)², for the appropriate hours from Table 6, measuring time prior to the expected time of maximum core temperature.

(ii) Enter in column 3 the Q-factors for $D = 0.02 \text{ cm}^2/\text{sec}$, $d = 2.41 \text{ in}$ from Table 1. As no values are tabulated for exactly $d = 2.41 \text{ in}$, enter those for the nearest tabulated diameter (2.50 in).

(iii) Enter in column 4 the sum of the entries of column 3 over consecutive constant values of column 2.

(iv) Enter in column 5 the product of the corresponding entries of columns 2 and 4.

(v) Enter at the foot of column 5 the value of A , the sum of the individual entries in column 5. Here $A = 0.715$.

(vi) Note the value of Y_0 , the entry in column 2 for time 0-1 hour. Here $Y_0 = 0.99$.

(vii) Calculate the value of cyclic rating factor from eqn. (6), namely

$$\text{Cyclic rating factor} = \frac{1}{\sqrt{[(1-k)Y_0 + kA]}} = 1.15$$

It is convenient to examine here the magnitude of the error introduced by utilizing the Q-factors of Table 1 for $d = 2.50 \text{ in}$ in place of $d = 2.41 \text{ in}$. If the calculation is repeated using the Q-factors corresponding to $d = 2.25 \text{ in}$, it will be found that the consequent value of A is 0.720. Clearly the error is trivial. It may also be mentioned that there is no significant difference between the values of k obtained when the depth of cable burial is referred to upper cable surface or upper cable axis.

Effect of Varying Soil Constants.

It is also possible to examine the effect on the cyclic rating factor of variation of thermal resistivity and diffusivity of the soil. The quantity k depends only on the soil thermal resistivity, and appropriate values of k for any soil thermal resistivity may be calculated with the use of eqn. (9). The quantity A depends on soil thermal diffusivity alone: only two values are covered, namely 0.02 and 0.01 cm^2/sec by Tables 1 and 2, respectively.

Calculations made for the cable and load curve of the present example for a range of soil thermal characteristics give the following cyclic rating factors:

$D \backslash g$	60	90	120
0.01	1.16	1.17	1.18
0.02	1.13	1.14	1.15

The cyclic rating factors increase with increasing soil thermal resistivity and decreasing soil thermal diffusivity.

(5.2) Example 2—Isolated Ducted Circuit, Various Soil Constants, Effect of Depth

Cable under consideration.—3-core screened 0.50in², 33 kV p.i.l.s. cable, armoured and served with added sheath protection, conforming to B.S. 480: 1954.

Method of laying.—In single-way duct at depth 36in of duct axis below ground surface.

Soil constants.—Soil thermal diffusivity $D = 0.02 \text{ cm}^2/\text{sec}$. Soil thermal resistivity $g = 120^\circ \text{C-cm/watt}$.

Load.—As described above, illustrated in Fig. 2 and tabulated in Table 6.

Additional data required.—Inner duct diameter = 6.00 in. (From B.S. 480: 1954 and E.R.A. Report Ref. F/T183.) For this cable at the above depth of burial and soil thermal resistivity, k is equal to 0.51 (from Table 5).*

Method of Calculation and Effects of Varying Soil Constants and Depth of Laying.

The calculation is made as described in detail in Example 1, except that, for the entry in column 3, Q-factors from Table 1 for the inner duct diameter $d = 6.00 \text{ in}$ are utilized.

The resultant cyclic rating factor, again calculated from eqn. (6), is 1.11, while for the same range of soil thermal characteristics as in Example 1 we obtain the following cyclic rating factors:

$D \backslash g$	60	90	120
0.01	1.08	1.10	1.12
0.02	1.07	1.09	1.11

As indicated earlier in the paper a cyclic rating factor for non-standard depth of burial 4.24ft and for soil thermal diffusivity

* For two identically loaded cables in 2 single way-ducts, k as given by eqn. (11) is used in place of k .

$D = 0.02 \text{ cm}^2/\text{sec}$ may be obtained by utilizing Q -factors from Table 2 for cable diameter $d = 6/\sqrt{2} = 4.24 \text{ in}$ in the calculation of A . In this way a 1% increase in cyclic rating factor is found for the non-standard depth of burial 4.24 ft as compared with the depth of burial 3 ft. This result is valid for soil thermal resistivities $g = 60, 90$ and $120^\circ \text{C-cm/watt}$, and soil thermal diffusivity $0.02 \text{ cm}^2/\text{sec}$.

(6) SIMPLIFICATION OF CYCLIC RATING FACTOR CALCULATION FOR (SINGLE) PEAKED-TYPE CYCLIC LOAD

The method becomes even more simple when applied to the calculation of cyclic rating factors for peaked-type load cycles [see Fig. 1(c)]. Let such a peaked load cycle consist of a peak load persisting for x hours and a constant proportion p of this peak load persisting for the remaining $(24 - x)$ hours of the cycle. The square of the applied load cycle as a fraction of its peak value is then as illustrated in Fig. 1(c). The temperature at the end of any step is that due to the cable continuously loaded at peak unit squared load less the temperature at the end of the step due to the squared load $(1 - p^2)$ applied for $(24 - x)$ hours daily. If the accumulated sums of the entries of each column

of Tables 1 and 2 starting from $T_2 = 23$ are tabulated as a_x (as in the skeleton Tables 8 and 9), the cyclic rating factor for the load in question is

$$1/\sqrt{[1 - ka_x(1 - p^2)]} \quad (12)$$

where x is the duration of the peak load and a_x is given directly by Tables 8 and 9 for soil thermal diffusivities of 0.02 and $0.01 \text{ cm}^2/\text{sec}$, respectively, and for the same range of cable diameters as in Tables 1 and 2. In particular, for a rectangular load cycle $p = 0$ [Fig. 1(a)].

Tables 8 and 9 were initially obtained by summation from Tables 1 and 2 when the latter were obtained to four decimal places, and all four tables were subsequently rounded off to three decimal places. Hence there exist apparent discrepancies in the last digit between corresponding tables.

Shanklin and Buller have also given a simple method for the calculation of cyclic rating factors when the applied cyclic load is rectangular in form.⁸ However, as most British cables are small in diameter, with a consequent low thermal capacity, the minimum loss factor of 0.3 considered in the paper could be usefully reduced while still not necessitating the introduction of

Table 8
VALUES OF a_x FOR SOIL THERMAL DIFFUSIVITY $0.02 \text{ cm}^2/\text{SEC}$

x	Cable diameter or inner duct diameter (in)												x
	0.50	0.75	1.00	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	6.00	
1	0.461	0.498	0.526	0.570	0.609	0.642	0.670	0.695	0.720	0.742	0.762	0.801	1
2	0.391	0.422	0.446	0.483	0.516	0.544	0.569	0.590	0.612	0.632	0.650	0.686	2
3	0.346	0.373	0.394	0.427	0.456	0.482	0.504	0.522	0.542	0.560	0.576	0.609	3
4	0.311	0.336	0.355	0.385	0.410	0.434	0.453	0.470	0.488	0.505	0.519	0.549	4
5	0.283	0.305	0.322	0.349	0.373	0.394	0.411	0.427	0.443	0.459	0.472	0.499	5
6	0.258	0.278	0.294	0.318	0.340	0.359	0.376	0.389	0.404	0.418	0.430	0.456	6
7	0.236	0.254	0.269	0.291	0.310	0.329	0.343	0.356	0.369	0.382	0.393	0.417	7
8	0.216	0.233	0.246	0.266	0.283	0.300	0.314	0.325	0.337	0.350	0.359	0.381	8
10	0.180	0.194	0.205	0.221	0.236	0.250	0.261	0.270	0.280	0.291	0.299	0.317	10
12	0.148	0.159	0.168	0.182	0.193	0.205	0.214	0.221	0.230	0.239	0.245	0.260	12
14	0.119	0.128	0.135	0.146	0.155	0.165	0.172	0.177	0.184	0.192	0.197	0.209	14
16	0.092	0.099	0.105	0.113	0.120	0.128	0.133	0.137	0.142	0.148	0.152	0.162	16
18	0.067	0.072	0.076	0.082	0.087	0.093	0.097	0.100	0.104	0.108	0.111	0.118	18
20	0.044	0.047	0.049	0.053	0.057	0.060	0.063	0.064	0.067	0.070	0.072	0.076	20
22	0.021	0.023	0.024	0.026	0.027	0.029	0.031	0.031	0.033	0.034	0.035	0.037	22

Table 9
VALUES OF a_x FOR SOIL THERMAL DIFFUSIVITY $0.01 \text{ cm}^2/\text{SEC}$

x	Cable diameter or inner duct diameter (in)												x
	0.50	0.75	1.00	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	6.00	
1	0.519	0.560	0.592	0.642	0.685	0.721	0.752	0.781	0.805	0.826	0.846	0.882	1
2	0.446	0.482	0.509	0.552	0.590	0.621	0.649	0.675	0.697	0.718	0.737	0.775	2
3	0.399	0.430	0.455	0.493	0.527	0.555	0.580	0.604	0.625	0.644	0.662	0.698	3
4	0.362	0.390	0.413	0.447	0.478	0.504	0.527	0.549	0.568	0.585	0.602	0.635	4
5	0.331	0.357	0.377	0.408	0.437	0.461	0.481	0.502	0.519	0.535	0.551	0.581	5
6	0.303	0.327	0.346	0.374	0.401	0.422	0.441	0.460	0.476	0.491	0.505	0.534	6
7	0.278	0.301	0.318	0.344	0.368	0.388	0.405	0.423	0.437	0.451	0.464	0.491	7
8	0.255	0.276	0.292	0.316	0.338	0.356	0.372	0.388	0.402	0.414	0.426	0.451	8
10	0.214	0.231	0.244	0.265	0.283	0.299	0.312	0.325	0.337	0.347	0.358	0.378	10
12	0.177	0.191	0.202	0.219	0.234	0.247	0.258	0.269	0.279	0.287	0.296	0.313	12
14	0.143	0.154	0.163	0.177	0.189	0.200	0.209	0.217	0.225	0.232	0.239	0.253	14
16	0.111	0.120	0.127	0.137	0.147	0.155	0.162	0.169	0.175	0.181	0.186	0.197	16
18	0.082	0.088	0.093	0.100	0.108	0.114	0.118	0.124	0.128	0.132	0.136	0.144	18
20	0.053	0.057	0.060	0.065	0.070	0.074	0.077	0.080	0.083	0.086	0.089	0.094	20
22	0.026	0.028	0.030	0.032	0.034	0.036	0.038	0.039	0.041	0.042	0.044	0.046	22

the cable thermal capacity. A comparison between the cyclic rating factors obtained by Shanklin and Buller's method and by the present method is given for an individual case, but is preceded by a further example illustrating the present method.

(6.1) Example 3—Isolated Circuit Laid Direct, (Single) Peaked-Type Load, Standard Conditions of Laying, Various Soil Constants

Cable under consideration.—3-core belted 0.30 in² 22 kV p.i.l.s. cable, armoured and served with added sheath protection, conforming to B.S. 480: 1954.

Method of laying.—Direct in the ground at depth 36 in of cable axis below ground surface.

Soil constants.—Soil thermal diffusivity $D = 0.02 \text{ cm}^2/\text{sec}$. Soil thermal resistivity $g = 120^\circ \text{C-cm/watt}$.

Load.—Peaked load of 100% current for 6 hours and 60% current for 18 hours.

Additional data required.—Overall diameter of cable = 3.04 in (from B.S. 480: 1954). For this cable at the above depth of burial and soil thermal resistivity k is equal to 0.68 (from Table 4).

Method of Calculation.

As 100% current is applied for 6 hours, read off a_x for $x = 6$ and $d = 3.00$ from Table 8. Here $a_x = 0.376$. As 60% current is applied for the remaining 18 hours, $p = 0.6$. The cyclic rating factor is then calculated directly from eqn. (12) and is

$$1/\sqrt{(1 - 0.68 \times 0.376 \times 0.64)} = 1.09$$

Effect of Varying Soil Constants.

For the same range of soil thermal characteristics as in Examples 1 and 2 above, the following cyclic rating factors result:

$D \backslash g$	60	90	120
0.01	1.08	1.10	1.11
0.02	1.07	1.08	1.09

(7) COMPARISON OF CYCLIC RATING FACTORS CALCULATED BY ALTERNATIVE METHODS FOR (SINGLE) PEAKED-TYPE CYCLIC LOAD

Consider rectangular load cycles of loss factors 0.3 and 0.5 applied to the same cable as in Example 3 for soil thermal resistivity and diffusivity of $80^\circ \text{C-cm/watt}$ and $0.01 \text{ cm}^2/\text{sec}$ respectively. The consequent attainment factors (ratio of maximum core temperature rise under rectangular load cycle to maximum core temperature rise under equal amplitude steady load) may readily be evaluated by Shanklin and Buller's method,⁸ utilizing Table 2 of that paper, and by the method outlined in the present paper, so that a direct numerical comparison will be obtained for an individual case.

The value of k for $g = 80$ is 0.585, while loss factors 0.3 and 0.5 for rectangular load cycles correspond to $x = 7.2$ and $x = 12$ respectively, so that from Table 9 for cable diameter 3.04 in the corresponding values of a_x are 0.399 and 0.259. Hence the attainment factors obtained by the method of the present paper are given by $1 - ka_x$, namely 0.77 and 0.85 respectively for loss factors of 0.3 and 0.5.

For the application of Shanklin and Buller's method we need $R_{int} = (G'/3) + G_s$ in the present notation. With a dielectric thermal resistivity of $550^\circ \text{C-cm/watt}$ and a protective covering thermal resistivity of $500^\circ \text{C-cm/watt}$ as used in the derivation of the tabulated values of k , R_{int} is found to be 1.18 thermal

ohms/ft, while a soil thermal diffusivity of $0.01 \text{ cm}^2/\text{sec}$ is $5.6 \text{ in}^2/\text{h}$ in the required units. Hence the parameter $r/2R_{int}\sqrt{k}$ of Shanklin and Buller's paper, where k is the soil thermal diffusivity in square inches per hour and r is the cable radius in inches, assumes the value 0.273 and the corresponding attainment factors for loss factors of 0.3 and 0.5 are 0.75 and 0.84, respectively, compared with 0.77 and 0.85 obtained above by the method outlined in the present paper. This represents fairly close agreement between the results of the two methods for the present example, the agreement being even closer when the consequent cyclic rating factors are compared.

(8) CHOICE OF APPROPRIATE VALUES FOR THERMAL DIFFUSIVITY AND RESISTIVITY OF THE SOIL

The soil thermal diffusivity D and resistivity g vary substantially along the length of a cable at a given time and vary with time at a given point of the cable. From the above examples it will be observed that the variation of cyclic rating factors with soil thermal diffusivity for typical applied load cycles is extremely small. As, further, the assumed values for this diffusivity in the paper are overestimates rather than underestimates, the consequent cyclic rating factors will be on the safe side, so that a lack of knowledge of soil thermal diffusivity appears to be no obstacle in the utilization of the method. The variation of cyclic rating factors with soil thermal resistivity is somewhat more important, but by calculating the cyclic rating factor for an estimated minimum value of soil thermal resistivity, a safe cyclic rating factor is obtained. It should be stressed that the cyclic rating factors of this paper are to be applied to the current rating of cables buried in ground of the appropriate soil thermal resistivity. Current ratings for cables conforming to B.S. 480: 1954 are given in an E.R.A. report,¹⁹ which also contains correction factors for variation of soil thermal resistivity.

(9) ACKNOWLEDGMENTS

The work described in the paper was carried out under the direction of Mr. L. Gosland, whose assistance, particularly in Section 4 of the paper, and whose encouragement, the author gratefully acknowledges. The author is also indebted to Miss Sheila Johnson for carrying out the lengthy calculations involved, and to the Joint Acting Directors of the Electrical Research Association, for permission to publish the paper.

(10) REFERENCES

- (1) 'Tables of Current Ratings of British Standard Cables', E.R.A. Technical Report Ref. F/T128 (1939). (Based on cables conforming with British Standards B.S. 480: 1933 and B.S. 608: 1935.)
- (2) WHITEHEAD, S., and HUTCHINGS, E. E.: 'Current Rating of Cables for Transmission and Distribution', *Journal I.E.E.*, 1938, **83**, p. 517.
- (3) WHITEHEAD, S., and HUTCHINGS, E. E.: 'Current Ratings of Cables for Transmission and Distribution', E.R.A. Technical Report Ref. F/T131, 1939. (Containing several amendments to Ref. 2 above.)
- (4) BULLER, F. H.: 'Thermal Transients on Buried Cables', *Transactions of the American I.E.E.*, 1951, **70**, Part I, p. 45.
- (5) KEATING, T. T.: 'The Cyclic Heating of "H" Type and "S.L." Type Buried Power Cables', *Journal of the Institution of Engineers, Australia*, 1950, **22**, p. 149.
- (6) NEHER, J. H.: 'The Determination of Temperature Transients in Cable Systems by means of an Analogue Computer', *Transactions of the American I.E.E.*, 1951, **70**, Part II, p. 1361.

- (7) 'The Effect of Loss Factor on the Temperature Rise of Pipe Cable and Buried Cables', American I.E.E. Committee Report, *ibid.*, 1953, 72, Part III, p. 530.
- (8) SHANKLIN, G. B., and BULLER, F. H.: 'Cyclic Loading of Buried Cable and Pipe Cable', *ibid.*, 1953, 72, Part III, p. 535.
- (9) NEHER, J. H.: 'Procedures for Calculating the Temperature Rise of Pipe Cable and Buried Cables for Sinusoidal and Rectangular Loss Cycles', *ibid.*, 1953, 72, Part III, p. 541.
- (10) WISEMAN, R. J.: 'An Empirical Method for Determining Transient Temperatures of Buried Cable Systems', *ibid.*, 1953, 72, Part III, p. 545.
- (11) Discussion on Refs. 6 to 9 above, *ibid.*, 1953, 72, Part III, p. 551.
- (12) MELSOM, S. W., and FAWSETT, E.: 'Permissible Current Loading of British Standard Impregnated Paper Insulated Electric Cable', *Journal I.E.E.*, 1923, 61, p. 517.
- (13) 'Current Carrying Capacity of Impregnated Paper, Rubber and Varnished Cambric Insulated Cables', Publication No. P-29-226. (Insulated Power Cable Engineers Association, New York, 1943.)
- (14) British Standard 480: 1954.
- (15) CARSLAW, H. S., and JAEGER, J. C.: 'Conduction of Heat in Solids' (Oxford University Press, 1947). (a) p. 221, (b) pp. 283-4.
- (16) 'Supplement to Report, Ref. F/T12, on the Heating of Buried Cables', E.R.A. Technical Report Ref. F/T15, 1925.
- (17) Tables of Sine, Cosine and Exponential Integrals (National Bureau of Standards, 1940), 1.
- (18) 'British Electricity Authority, First Report and Statement of Accounts for the Period 13th August, 1947, to 31st March, 1949' (H.M. Stationery Office, London, 1949), p. 260.
- (19) 'Tables of Current Ratings for British Standard Paper Insulated Cables', E.R.A. Technical Report Ref. F/T183, 1955.

(11) APPENDICES

(11.1) Method of Calculation of Cyclic Rating Factors taking into Account the Thermal Capacity of the Cable

The transient core temperature rise $\theta_c(T)$ of a cable at time T hours after the application to the cable of a step-function load has been shown by Whitehead and Hutchings^{2,3} to be given approximately by

$$\frac{\theta_c(T)}{\theta_c(\infty)} = (1 - k)[1 - \exp(-3600T/SG')] + k\alpha(T)$$

$$= (1 - k)\beta(T) + k\alpha(T)$$

where $\beta(T) = 1 - \exp(-3600T/SG')$

k = Ratio of outer cable temperature rise to core temperature rise under steady conditions.

S = Thermal capacity of the cable, namely the sum of the thermal capacity of the loaded conductors and one half the thermal capacity of the remainder of the cable.

G' = Internal thermal resistance of the cable.

$\alpha(T)$ = Ratio of transient outer cable temperature rise to steady-state outer cable temperature rise.

From the above, the transient core temperature rise of a cable at time T_2 hours after the application to the cable of a rectangular pulse of duration $T_1 - T_2$ hours is given as a ratio of the steady-

state core temperature rise $\theta_c(\infty)$ on application of an equal-amplitude step-function load by

$$\frac{\theta_c(T_1) - \theta_c(T_2)}{\theta_c(\infty)}$$

$$= (1 - k)[\beta(T_1) - \beta(T_2)] + k[\alpha(T_1) - \alpha(T_2)]$$

$$= (1 - k)[\exp(-3600T_2/SG') - \exp(-3600T_1/SG')] + k[\alpha(T_1) - \alpha(T_2)]$$

Hence the transient core temperature rise of a buried cable at time T_2 hours after the application to the cable of a very large number of identical rectangular load pulses, each of duration $T_1 - T_2$ hours, and each separated by equal intervals of $24 - (T_1 - T_2)$ hours, has an upper limit given as a ratio of the steady-state core temperature rise $\theta_c(\infty)$ by

$$\frac{\sum_{m=0}^{\infty} [\theta_c(T_1 + 24m) - \theta_c(T_2 + 24m)]}{\theta_c(\infty)}$$

$$= (1 - k) \sum_{m=0}^{\infty} [\beta(T_1 + 24m) - \beta(T_2 + 24m)]$$

$$+ k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)]$$

$$= (1 - k) \sum_{m=0}^{\infty} \exp[-3600(T_2 + 24m)/SG']$$

$$- \exp[-3600(T_1 + 24m)/SG']$$

$$+ k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)]$$

$$= \frac{(1 - k)[\exp(-3600T_2/SG') - \exp(-3600T_1/SG')]}{1 - \exp(-24 \times 3600/SG')}$$

$$+ k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)]$$

For all British Standard cables the term $\exp(-24 \times 3600/SG')$ is much less than unity and may be neglected, so that the above expression becomes

$$(1 - k)[\exp(-3600T_2/SG') - \exp(-3600T_1/SG')] + k \sum_{m=0}^{\infty} [\alpha(T_1 + 24m) - \alpha(T_2 + 24m)]$$

In particular, for pulses of 1 h duration, so that $T_1 - T_2 = 1$, and for $T_2 = 0(1)23$, the sum of the infinite series is merely the Q-factor given by eqn. (5) and tabulated in Tables 1 and 2. (A more detailed description of these tables has already been given.) For any individual cable it is a relatively simple matter to calculate the values of

$$[\exp(-3600T_2/SG') - \exp(-3600T_1/SG')]$$

for $T_1 - T_2 = 1$, $T_2 = 0(1)23$, and to enter these values in an additional column of a calculation sheet such as that presented in Table 7. The sum of the products of these values and the corresponding values of column 2, the equivalent squared applied load, would then be calculated. Denoting this result by B , the new cyclic rating factor would be given by

$$\frac{1}{\sqrt{[(1 - k)B + kA]}}$$

with A obtained as in Table 7.

When the form of the applied load in the neighbourhood of its

peak is such that an initial half-hourly representation is markedly preferable to an hourly representation, it becomes important to split the values of Tables 1 and 2 for $T_2 = 0$ into those for two half-hourly periods. This has been carried out and the consequent values are given in Table 3. When this table is utilized, consideration may also be given to the effect of the cable thermal capacity on the cyclic rating factor.

(11.2) Detailed Comparison of Formulae for $\alpha(t)$

Two different formulae for $\alpha(t)$, the ratio of transient to steady-state outer cable temperature rise on application to a cable of a step-function load, are given in eqns. (2) and (3) of the paper. To investigate the percentage difference between these formulae for the range of parameters involved, we utilize the well-known expansion for the exponential integral of a small negative argument,¹⁷

$$-Ei(-y) = -\gamma - \log_e y + y - \frac{y^2}{4} + \dots$$

where $\gamma = 0.5772 \dots$ = Euler's constant. Denoting $x = d^2/16Dt$, eqns. (2) and (3) become respectively,

$$\alpha(t) = \frac{-Ei(-x) + Ei(-l^2/Dt)}{\log_e(16l^2/d^2)} \quad (2a)$$

$$\alpha(t) = \frac{-Ei(-2x) + (\log_e 2) \exp(-2x) + Ei(-l^2/Dt)}{(\log_e 16l^2/d^2)} \quad (3a)$$

The denominators of these two expressions for $\alpha(t)$ are the same, as is the term $Ei(-l^2/Dt)$ in the numerator. Hence a

comparison is required between the values of $-Ei(-x)$ and $-Ei(-2x) + (\log_e 2) \exp(-2x)$. For British Standard cables and soil thermal diffusivity, x is a very small quantity for times t greater than or equal to one hour. Using the above expansion for the exponential integral we have

$$-Ei(-x) = -\gamma - \log_e x + x - \frac{x^2}{4} \dots \quad (2b)$$

$$-Ei(-2x) + (\log_e 2) \exp(-2x) = -\gamma - (\log_e 2x)$$

$$+ 2x - \frac{(2x)^2}{4} + \dots + (\log_e 2) \left[1 - 2x + \frac{(2x)^2}{2} \dots \right]$$

$$= -\gamma - \log_e x + 2x(1 - \log_e 2) + x^2(2 \log_e 2 - 1) \quad (3b)$$

For sufficiently small values of x , eqns. (2b) and (3b) yield substantially the same result. Further, the maximum percentage difference between eqns. (2a) and (3a) occurs when x is largest, which in the present context is for time $t = 1$ h, and for soil thermal diffusivity $D = 0.01 \text{ cm}^2/\text{sec}$. Further when $t = 1$ h the term $Ei(-l^2/Dt)$ is negligible in the numerator so that the maximum percentage difference between eqns. (2a) and (3a) is the same as that between eqns. (2b) and (3b). Taking one of the largest single cable diameters of B.S. 480: 1954, namely $d = 4.50$ in, and with $D = 0.01 \text{ cm}^2/\text{sec}$, $t = 1 \text{ h} = 3600 \text{ sec}$, then $x = 0.226$ and there is a 5% difference between eqns. (2b) and (3b). (For $t = 2$ h this is reduced to 2%.) The consequent difference in the resultant cyclic rating factors for this unfavourable case would be much below 5%.

THE PRODUCTION OF A SINUSOIDAL FLUX WAVE, WITH PARTICULAR REFERENCE TO THE INDUCTOR ALTERNATOR

By N. N. HANCOCK, B.Sc.(Eng.), M.Sc.Tech., Associate Member.

(The paper was first received 26th April, and in revised form 25th June, 1956. It was published as an INSTITUTION MONOGRAPH in October, 1956.)

SUMMARY

After a brief reference to the impossibility of producing a sinusoidal flux-wave with the more common forms of synchronous generator, the paper deals with the conditions necessary for a sinusoidal flux-variation in inductor-type alternators.

LIST OF SYMBOLS

- B_m = Mean gap flux density measured at the armature surface, webers/m².
 B_c = Peak value of the sinusoidal component of the flux density, webers/m².
 F, G, H = Functions.
 M = Magnetic potential, ampere-turns.
 P, Q = Constants.
 R = Radius of armature bore.
 $a = \mu_0 \pi M / B_c$, equal to $\mu_0 p M / R B_c$ in the circular form.
 $c = B_m / B_c$.
 f, g, h = Functions.
 l_g = Radial air-gap at the centre-line of the rotor teeth, m.
 p = Number of pairs of poles = number of rotor teeth.
 r = Distance of a point from the centre of the armature bore, m.
 x = Abscissa of a point in the developed form, m.
 y = Distance of a point from the surface of the armature bore, m.
 α, β = Constants.
 θ = Polar co-ordinate in terms of electrical angles, rad.
 Θ = Polar co-ordinate in terms of mechanical angles, rad.
 τ = Pole pitch measured on the armature surface = half the distance between adjacent rotor teeth measured on the armature surface, m.
 $\phi = \pi y / \tau$ in the developed form, $\log_e (R/r)^p$ in the circular form.
 $0, \pi$ (as subscripts) indicate values when $\theta = 0$ and $\theta = \pi$ respectively.

(1) INTRODUCTION

The desirability of eliminating harmonics in the voltages and currents of a.c. systems has long been recognized, and considerable attention has been devoted to the reduction of the harmonics in the open-circuit terminal voltage of the normal heteropolar generator. The technique is to rely on the armature winding arrangement to reduce to an acceptable level the harmonics which inevitably arise from the difficulty of producing a flux wave with sinusoidal distribution. The production of a sinusoidal flux wave presents different problems according to the type of rotor employed. With turbo-type rotors the problem is practically insoluble, since, apart from the differing effective permeance over the slotted and unslotted portions, a sinusoidal distribution of m.m.f. presents considerable manufacturing difficulties. Equally, as shown by Hague,¹ the pole profile required for a sinusoidal

flux wave in salient-pole machines is impracticable, leaving no space for field windings, producing considerable leakage flux, and providing no flux path between poles at the roots. These difficulties lead to the use of a cylindrical rotor with a 3-phase type of winding for generators used for high-voltage testing in which waveform is of particular importance. This produces the harmonics of a normal 3-phase winding, the phase relationships being those which occur at the instant of maximum current in one phase. These harmonics can be reduced by employing a 3-phase winding with a variety of turns per coil, such as that described by Krebs.² There is not even a theoretical possibility of producing a machine with the above types of rotor which has no harmonics in the flux wave.

It remains to consider the inductor type of generator. Such machines, obsolete for generation at normal supply frequency before detailed attention was given to waveform, were revived for supplies at frequencies in the 1000 to 10000 c/s range for induction heating and for supplies to servo systems. A common form of rotor tooth profile resembles that of a gear wheel tooth. Not only is this not the ideal shape, but also, in contrast with other types of generator, the inductor alternator requires for a sinusoidal flux wave a profile which is quite practicable. A profile of ideal form for an 8-pole rotor is shown in Fig. 1.

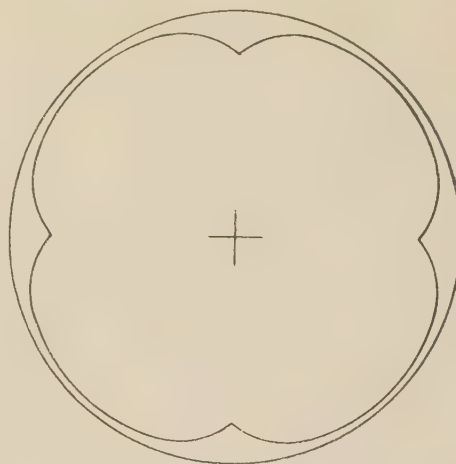


Fig. 1.—An ideal form for an eight-pole rotor.

Analysis leads to the equations of lines of constant magnetic potential which will produce the required distribution of flux density at the armature surface. Interest is confined in practice to one of these equipotentials, namely that of the rotor surface. The solution for the developed form first given by Walker,³ showing that there is a limiting optimum condition, is derived by direct solution of the differential equations in Section 7.1. In Section 7.2 this solution is extended to include the curvature of the armature, the effect of which is negligible in multipolar generators but may be of significance if the number of poles is

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Hancock is at the College of Science and Technology, University of Manchester.

small, as for example in a generator for normal supply frequency. The implications of these equations are considered with special reference to the optimum case and the ease of computation of practicable rotor tooth profiles.

(2) ANALYSIS

(2.1) Assumptions

The analysis, in common with that for other flux distributions in electrical machinery, is necessarily based on the assumption that the flux densities are such that the reluctance of the iron parts of the circuit is negligible compared with that of the air-gap. Fringing at the ends of the machine is neglected, so that the problem may be treated as two-dimensional.

It is assumed that the surface of the armature is smooth. This entails neglecting the effect of the armature slot openings, which may be more serious than in heteropolar machines, since the number of slots per pole is frequently only one, and the minimum radial air-gap is small so that the effect of the slot-opening is enhanced. The type of machine which has relatively few armature slots, but in which the armature teeth have small 'sub-teeth' of the same pitch as the rotor teeth is excluded, since it presents a different type of problem.

(2.2) Analysis of the Developed Form

Since the analysis which neglects the curvature of the armature can be extended readily to include it, and since also the development is sufficiently accurate when the number of poles is large, and requires less arithmetic, it is convenient to treat it first.

In Fig. 2 the variation of flux density is expressed as a cosine

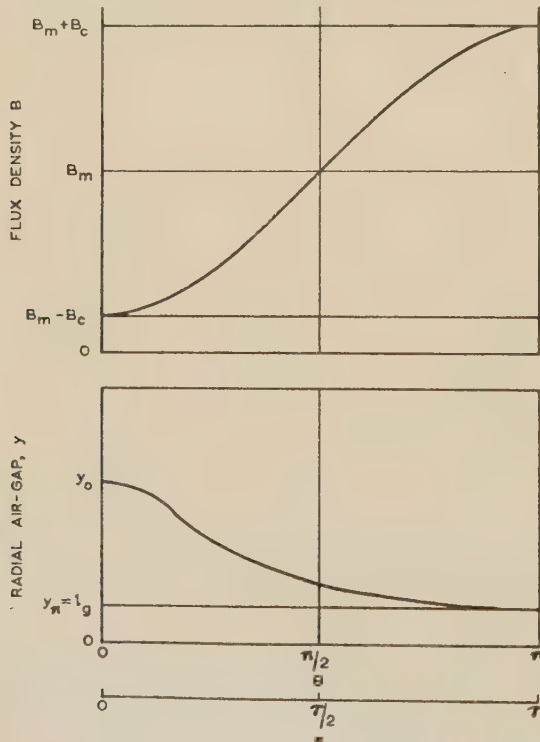


Fig. 2.—Variation of flux density B and radial air-gap y .

wave in order to permit one of the axes of symmetry to be taken as the zero of the angle θ , the axis midway between the rotor teeth being selected so that the data are as follows:

The radial flux density at the armature surface is

$$\mu_0 \left(\frac{\partial M}{\partial y} \right)_{y=0} = B_m - B_c \cos(\pi x / \tau) \quad (1)$$

The tangential flux density at the armature surface is zero:

$$\mu_0 \left(\frac{\partial M}{\partial x} \right)_{y=0} = 0 \quad (2)$$

The Laplace equation

$$\mu_0 \left(\frac{\partial^2 M}{\partial x^2} + \frac{\partial^2 M}{\partial y^2} \right) = 0 \quad (3)$$

applies for all values of x and y .

Lastly, the potential of the armature is arbitrarily taken as zero, so that

$$M = 0 \text{ when } y = 0 \quad (4)$$

Since $\pi x / \tau = \theta$, these equations may be rewritten in terms of θ :

$$\mu_0 \left(\frac{\partial M}{\partial y} \right)_{y=0} = B_m - B_c \cos \theta \quad (1)$$

$$\mu_0 \left(\frac{\partial M}{\partial x} \right)_{y=0} = 0 \quad (2)$$

$$\mu_0 \left[\left(\frac{\pi}{\tau} \right)^2 \left(\frac{\partial^2 M}{\partial \theta^2} \right) + \frac{\partial^2 M}{\partial y^2} \right] = 0 \quad (3)$$

$$M_{y=0} = 0 \quad (4)$$

It is shown in Section 7.1 that the solution of these equations is

$$\mu_0 M = B_m y - \frac{\tau}{2\pi} B_c (e^{\pi y / \tau} - e^{-\pi y / \tau}) \cos \theta \quad (5)$$

which can be written

$$\mu_0 M = B_m y - \frac{\tau}{\pi} B_c \sinh(\pi y / \tau) \cos \theta \quad (6)$$

or

$$a = c\phi - \sinh \phi \cos \theta \quad (7)$$

where $a = \mu_0 \pi M / \tau B_c$, $\phi = \pi y / \tau$ and $c = B_m / B_c$

(2.3) Analysis of Circular Form

The more exact case of the circular armature can be treated by the conformal transformation $w = e^z$, but, since this involves the idea of an infinite pole-pitch in the development, the direct method will be used.

Taking the flux as directed radially inwards from the armature surface, the data become in polar co-ordinates:

$$\mu_0 \left(\frac{\partial M}{\partial r} \right)_{r=R} = -B_m + B_c \cos p\Theta \quad (8)$$

$$\mu_0 \left(\frac{\partial M}{\partial \Theta} \right)_{r=R} = 0 \quad (9)$$

$$M_{r=R} = 0 \quad (10)$$

$$\mu_0 \left(\frac{\partial^2 M}{\partial r^2} + \frac{1}{r} \frac{\partial M}{\partial r} + \frac{1}{r^2} \frac{\partial^2 M}{\partial \Theta^2} \right) = 0 \quad (11)$$

in which R is the radius of the armature bore, p is the number of pole-pairs, which is equal to the number of rotor teeth, and Θ is the mechanical angle and is equal to θ/p , so that the polar co-ordinates of a point in the field are (r, Θ) .

The solution of this equation is shown in Section 7.2 to be

$$\mu_0 M = \frac{RB_m}{p} \log_e (R/r)^p - \frac{RB_c}{p} \sinh [\log_e (R/r)^p] \cos p\Theta \quad (12)$$

As before, this can be rearranged as

$$\mu_0 \frac{pM}{RB_c} = \frac{B_m}{B_c} \log_e (R/r)^p - \sinh [\log_e (R/r)^p] \cdot \cos p\theta$$

or, as before, $a = c\phi - \sinh \phi \cos \theta$ (13)

where now

$$a = \mu_0 pM/RB_c, c = B_m/B_c \text{ and } \phi = \log_e (R/r)^p$$

Since $2\pi R = 2p\tau$, $p/R = \pi/\tau$; a is therefore the same as in the developed case and only ϕ is changed.

(3) SIGNIFICANCE OF THE EQUATIONS

As eqns. (7) and (13) are identical, the same comments will apply to both the developed and circular forms, and it is merely necessary to note that, whereas in the developed form infinite values of ϕ correspond to infinite values of y , in the circular case they denote zero values of r ; both are of course impracticable.

The equation $a = c\phi - \sinh \phi \cos \theta$ has several interesting characteristics. When $\theta = \pi$, $\cos \theta = -1$ and the equation becomes $a = c\phi + \sinh \phi$. Writing this as $\sinh \phi = a - c\phi$ and solving graphically as in Fig. 3, it can be seen that there is

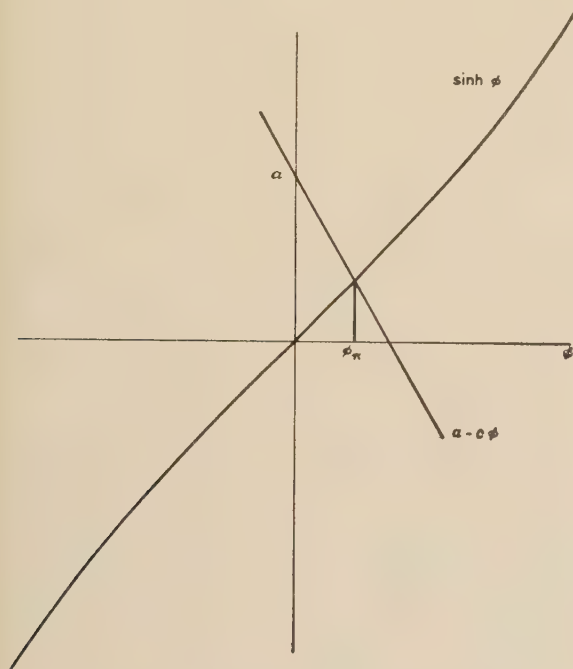


Fig. 3.—Graphical solution for ϕ_π .

only one value of ϕ_π for a given value of a , and hence only one value of y_π or r_π for a given potential.

However, when $\theta = 0$, $\cos \theta = 1$ so that the equation becomes $a = c\phi - \sinh \phi$. Solving this graphically as $\sinh \phi = c\phi - a$, it is seen in Fig. 4 that there may be two distinct positive real roots, two coincident positive roots, or no positive real roots, according to the values of a and c . There is also a negative root which is the reflection in the θ axis of the corresponding negative equipotential. The graphs of an equipotential corresponding to these three cases are shown in Fig. 5.

For a given distribution of flux to be obtained in practice, physical boundaries must be provided corresponding to two of the equipotential lines enclosing the space in which the distribution is required. In this case the two equipotentials are the surfaces of the rotor and the armature. It is necessary to

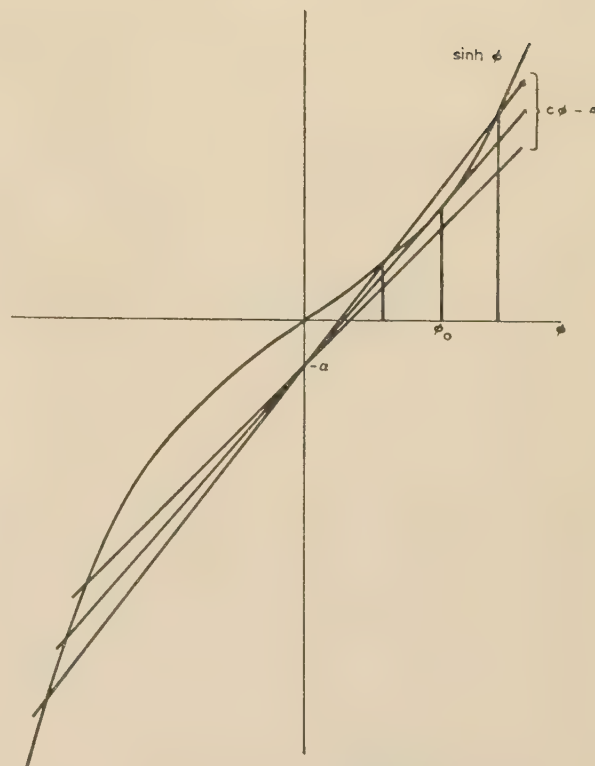


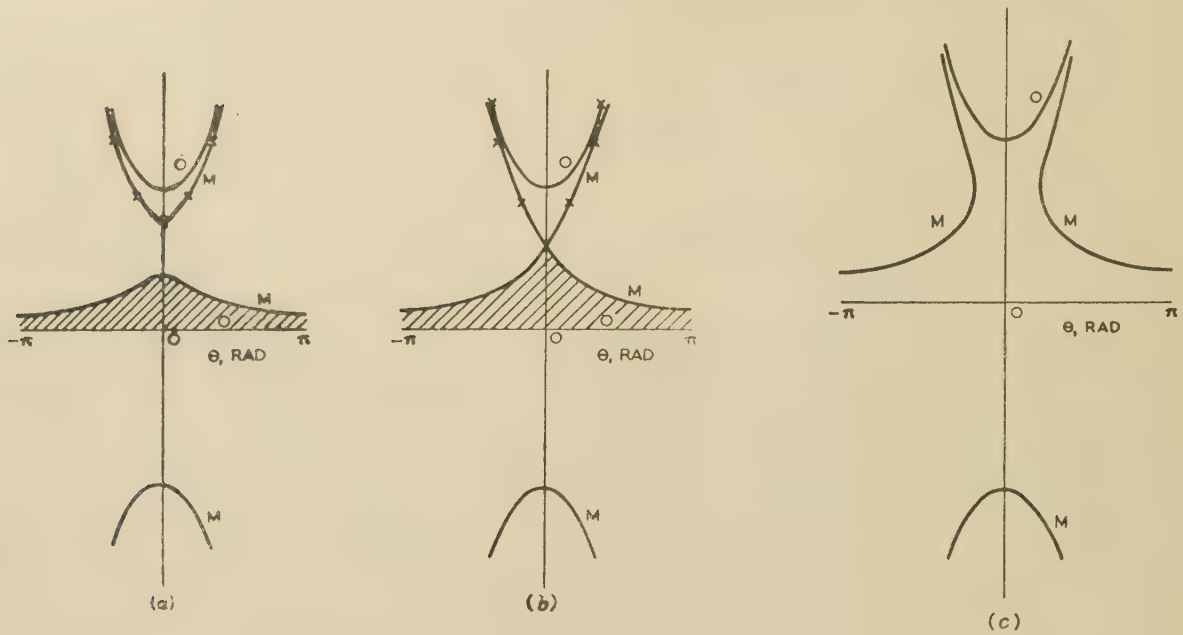
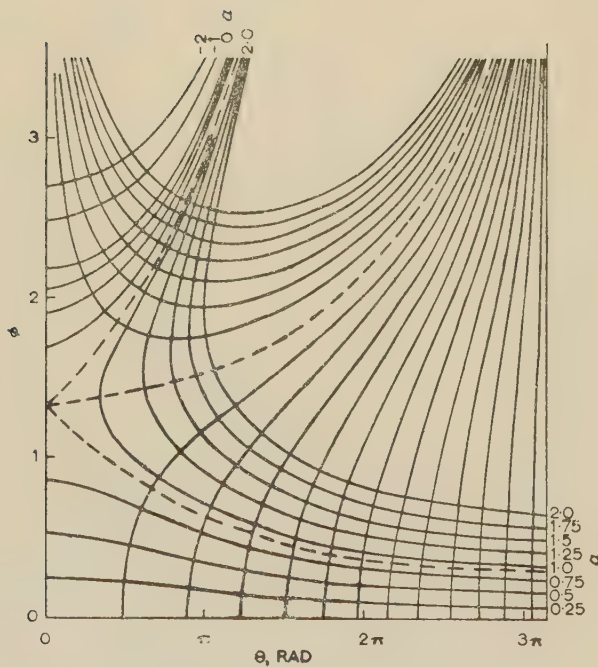
Fig. 4.—Graphical solution for ϕ_0 .

consider only a portion between the two axes of symmetry corresponding to $\theta = 0$ and $\theta = \pi$, since the pattern is merely repeated with alternate handings around the whole periphery. Now with the values of a and c giving equipotentials of the form shown in Fig. 5(a), the area shown shaded is completely enclosed by the zero-potential armature surface and the continuous part of the selected equipotential. The same applies to Fig. 5(b), except that there is now a cusp on the $\theta = 0$ axis. With Fig. 5(c), however, there is no such enclosure, the equipotential going off to infinite values of ϕ with $\theta = \pm \pi/2$ as asymptotes.

Although it is apparent that this cannot be reproduced in practice, there is a further difficulty not initially apparent. The armature surface is at zero potential, but it is not the complete zero-potential surface; in all cases there is an additional part as shown in Fig. 5. In Figs. 5(a) and 5(b) this part is external to the enclosed air-gap space and is of no practical importance. In Fig. 5(c), however, it is not external to the enclosed space and it, or some other adjacent equipotential surface, must be included between the equipotentials given by a and c . Fig. 6, which shows a number of equipotentials and flux lines, illustrates this more clearly. The significance of this in practice is that it makes it even more inaccurate to use a large finite value of ϕ as an approximation where an infinite value is indicated. It means also that in every inductor type of flux pattern there lurks a heteropolar pattern involving equipotentials corresponding to potentials M , 0 and $-M$. In fact, if c be made equal to zero, the solution reduces to that given by Hague¹ for the heteropolar machine.

(3.1) Optimum Conditions

It is apparent that since the maximum flux density ($B_m + B_c$) is limited, the ratio $c = B_m/B_c$ must be as small as possible in order that B_c shall be as large as possible. Since there is a practical lower limit to this ratio, this limiting case must be the

Fig. 5.—Zero- and M -equipotentials.Fig. 6.—Equipotentials and flux lines (for $c = 2$).

ideal, and it is necessary to determine the corresponding values of c and a .

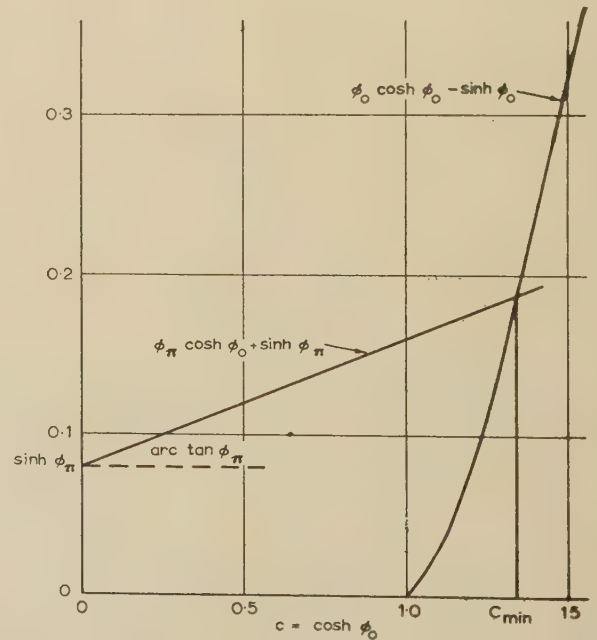
For a given radial air-gap l_g at the centre-line of the rotor teeth, ϕ_π is determined either as $\pi l_g / \tau$ or as $\log_e [R/(R - l_g)]^p$.

Hence $a = c\phi_\pi + \sinh \phi_\pi$ (14)

since θ is then equal to π .

When $\theta = 0$, the equation becomes

$$a = c\phi_0 - \sinh \phi_0 \quad . \quad . \quad . \quad (15)$$

Fig. 7.—Graphical solution for c_{min} .

Eliminating a from eqns. (14) and (15)

$$c = \frac{\sinh \phi_0 + \sinh \phi_\pi}{\phi_0 - \phi_\pi} \quad . \quad . \quad . \quad (16)$$

and differentiating eqn. (16),

$$\frac{dc}{d\phi_0} = \frac{(\phi_0 - \phi_\pi) \cosh \phi_0 - \sinh \phi_0 - \sinh \phi_\pi}{(\phi_0 - \phi_\pi)^2}$$

The minimum value of c thus occurs when

$$(\phi_0 - \phi_\pi) \cosh \phi_0 - \sinh \phi_0 - \sinh \phi_\pi = 0 \quad . \quad . \quad (17)$$

This can be solved most conveniently graphically by plotting $(\phi \cosh \phi - \sinh \phi)$ and $(\phi_\pi \cosh \phi + \sinh \phi_\pi)$ against $\cosh \phi$ as in Fig. 7, which, for any given case, involves drawing a single straight line to intersect a curve which can be plotted once for all.

Substituting for ϕ_0 from eqn. (17) in eqn. (16) it is seen that $c_{min} = \cosh \phi_0$ and can hence be read directly from the graph. The corresponding value of $\phi_0 = \text{arc cosh } c_{min}$ determines the radial air-gap midway between the rotor teeth and can be found directly if required.

If eqn. (17) be rewritten in terms of c_{min} instead of in terms of ϕ_0 it becomes

$$c_{min} (\text{arc cosh } c_{min} - \phi_\pi) - \sqrt{(c_{min}^2 - 1)} - \sinh \phi_\pi = 0$$

showing that it is the value of ϕ_π alone which determines the minimum value of c and hence the corresponding value of a . The possible values of c are thus restricted by the ratio of the air-gap at the centre of the rotor teeth to the pole-pitch, since these quantities determine ϕ_π . The limiting values are plotted in Fig. 8, which therefore shows the best value of c that can be obtained for a given pole-pitch and minimum air-gap.

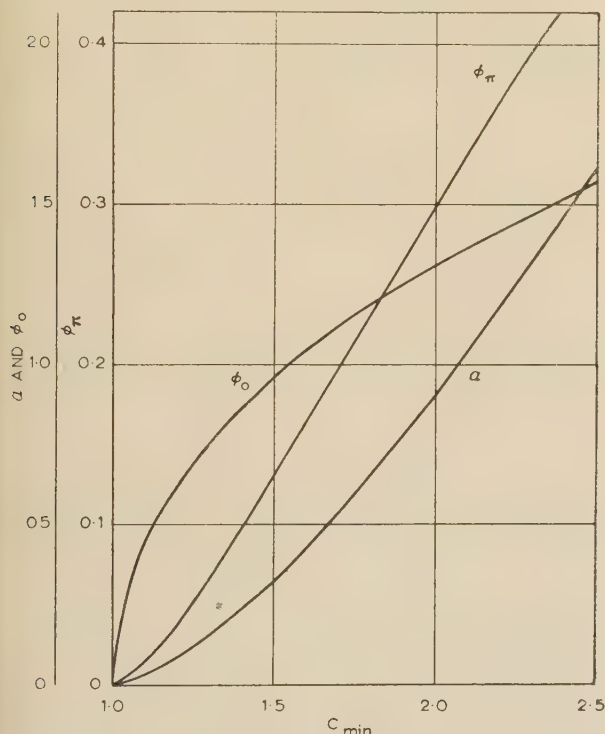


Fig. 8.—Relationship between ϕ_π , c_{min} , ϕ_0 and a .

When the values of a and c have been selected, whether or not the minimum value of c is used, it is necessary to find the values of ϕ for values of θ from 0 to π . The values of r can then be found and the rotor tooth profile plotted. Where the number of poles is large, as in the case of high-frequency generators, it is unnecessary to use the logarithmic expression, the relationship $\phi = \phi/\pi$ of the developed form being sufficiently accurate.

Fig. 9 shows the ϕ/θ curves for minimum values of c equal to 1.5, 2, 2.5, and 3. Since y is approximately proportional to ϕ they give an indication of the profile required.

(4) COMPUTATION

It is impracticable to solve for ϕ in terms of θ , and the calculation has to be in the form of solving for θ for selected values of ϕ between ϕ_π and ϕ_0 . Great accuracy is required in the arith-

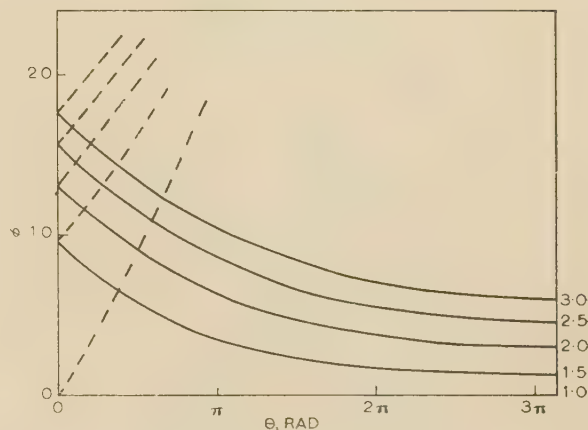


Fig. 9.—Optimum profiles.

metic, since differences are involved, and over a large part of the range a and $c\phi$ are of the same order of magnitude. Over the remainder of the range $\cos \theta$ approaches unity and a small error entails a considerably greater error in θ . Whilst expansion in series is attractive, it is impossible to generalize on its accuracy, since, although $y_\pi = l_g$ may be small compared with τ , y_0 may not be so.

(5) GENERAL COMMENTS

A slight increase of c above the minimum values causes a considerable decrease in y_0 and may therefore be justified if manufacture is thereby eased.

The analysis is based on the assumption that the rotor is exactly centred relative to the armature. Since the minimum air-gap is necessarily small for a reasonable value of c , a small eccentricity will produce a big change in the shape of the flux wave.

(5.1) Non-Sinusoidal Flux Waves

If a flux wave is expressed as $\mu_0 \left(\frac{\partial M}{\partial \theta} \right)_{r=R} = B_m - \sum B_n \cos n\theta$, the equations of the equipotentials are of the form

$$a = c\phi - \sum \frac{c_n}{n} \sinh n\phi \cos n\theta$$

It is therefore possible to design a tooth profile to produce any specified shape of flux wave. The numerical solution of the equation is tedious or difficult, leading to a quadratic when only the second harmonic is involved, a cubic with the third harmonic, and so on. The interest in this lies, however, in the fact that smaller values of c become practicable in some cases, and it may be preferable to design the profile for a known amount of harmonic of a specified order rather than to try approximating to the curve for a sinusoidal flux by using relatively small finite values of ϕ where large or infinite values are indicated. The harmonic may then be eliminated by the arrangement of the armature winding or by the use of filters.

(6) REFERENCES

- (1) HAGUE, B.: 'The Shape of Pole-Shoe required to produce a Sinusoidal Distribution of Air-Gap Flux Density', *Journal I.E.E.*, 1924, 62, p. 921.
- (2) KREBS: 'Winding Layouts for Dynamo-Electric Machines', British Patent No. 644676.
- (3) WALKER, J. H.: 'The Theory of the Inductor Alternator', *Journal I.E.E.*, 1942, Part II, 89, p. 238.

(7) APPENDICES

(7.1) Solution of Developed Form

$$\mu_0 \left(\frac{\partial M}{\partial y} \right)_{y=0} = B_m - B_c \cos \theta \quad (1)$$

$$\mu_0 \left(\frac{\partial M}{\partial \theta} \right)_{y=0} = 0 \quad (2)$$

$$\mu_0 \left[\left(\frac{\pi}{\tau} \right)^2 \left(\frac{\partial^2 M}{\partial \theta^2} \right) + \frac{\partial^2 M}{\partial y^2} \right] = 0 \quad (3)$$

$$M_{y=0} = 0 \quad (4)$$

From eqn. (1) it may be assumed that

$$\mu_0 \left(\frac{\partial M}{\partial y} \right) = \frac{df}{dy} - \frac{dg}{dy} \cos \theta$$

where f and g are functions of y such that $(df/dy)_{y=0} = B_m$ and $(dg/dy)_{y=0} = B_c$. Integrating with respect to y ,

$$\mu_0 M = f(y) - g(y) \cos \theta + h(\theta)$$

where h is a function of θ only. Differentiating with respect to θ ,

$$\mu_0 \left(\frac{\partial M}{\partial \theta} \right) = g(y) \sin \theta + dh/d\theta$$

so that
$$\mu_0 \left(\frac{\partial M}{\partial \theta} \right)_{y=0} = g(0) \sin \theta + dh/d\theta = 0$$

whence $dh/d\theta = -g(0) \sin \theta$. Integrating with respect to θ , $h = g(0) \cos \theta + \alpha$, where α is a constant. Hence $\mu_0 M = [f(y) + \alpha] - [g(y) - g(0)] \cos \theta$. Since $M_{y=0} = 0$, $\alpha = -f(0)$ and

$$\mu_0 M = [f(y) - f(0)] - [g(y) - g(0)] \cos \theta$$

Writing $F(y)$ for $[f(y) - f(0)]$ and $G(y)$ for $[g(y) - g(0)]$,

$$\mu_0 M = F(y) - G(y) \cos \theta$$

where $F(0) = 0$ and $G(0) = 0$.

Then

$$\mu_0 \left(\frac{\partial M}{\partial \theta} \right) = G(y) \sin \theta, \quad \mu_0 \left(\frac{\partial^2 M}{\partial \theta^2} \right) = G(y) \cos \theta$$

$$\mu_0 \left(\frac{\partial M}{\partial y} \right) = \frac{dF}{dy} - \frac{dG}{dy} \cos \theta, \quad \mu_0 \left(\frac{\partial^2 M}{\partial y^2} \right) = \frac{d^2 F}{dy^2} - \frac{d^2 G}{dy^2} \cos \theta$$

Substituting in eqn. (3),

$$\frac{d^2 F}{dy^2} - \frac{d^2 G}{dy^2} \cos \theta + \left(\frac{\pi}{\tau} \right)^2 G(y) \cos \theta = 0$$

for all values of θ and y . In particular when $\theta = \pi/2$ so that $\cos \theta = 0$, then $d^2 F/dy^2 = 0$ for all values of y ; therefore $F = \beta y + \gamma$, but since $F(0) = 0$, $\gamma = 0$, and since $dF/dy = df/dy = B_m$, $\beta = B_m$. Again, since $d^2 F/dy^2 = 0$,

$$\left[\frac{d^2 G}{dy^2} - \left(\frac{\pi}{\tau} \right)^2 G(y) \right] \cos \theta = 0$$

for all values of y and θ . Therefore $\frac{d^2 G}{dy^2} - \left(\frac{\pi}{\tau} \right)^2 G(y) = 0$ for all values of y . The solution of this second-order linear equation is $G = P e^{\pi y/\tau} + Q e^{-\pi y/\tau}$, where P and Q are constants. Since $G(0) = 0$, $P + Q = 0$, and since

$$(dG/dy)_{y=0} = (dg/dy)_{y=0} = B_c, \quad \frac{\pi}{\tau} (P - Q) = B_c$$

and

$$P = \frac{\tau B_c}{2\pi} \text{ and } Q = -\frac{\tau B_c}{2\pi}$$

Hence
$$\mu_0 M = B_m y - \frac{\tau}{2\pi} B_c (\epsilon^{\pi y/\tau} - \epsilon^{-\pi y/\tau}) \cos \theta \quad (5)$$

(7.2) Solution of Circular Form

$$\mu_0 \left(\frac{\partial M}{\partial r} \right)_{r=R} = -B_m + B_c \cos p\Theta \quad (8)$$

$$\mu_0 \left(\frac{\partial M}{\partial \Theta} \right)_{r=R} = 0 \quad (9)$$

$$M_{r=R} = 0 \quad (10)$$

$$\mu_0 \left(\frac{\partial^2 M}{\partial r^2} + \frac{1}{r} \frac{\partial M}{\partial r} + \frac{1}{r^2} \frac{\partial^2 M}{\partial \Theta^2} \right) = 0 \quad (11)$$

From eqn. (8) it may be assumed that

$$\mu_0 \left(\frac{\partial M}{\partial r} \right) = \frac{df}{dr} + \frac{dg}{dr} \cos p\Theta$$

where f and g are functions of r only, such that $(df/dr)_{r=R} = -B_m$ and $(dg/dr)_{r=R} = B_c$. Integrating with respect to r ,

$$\mu_0 M = f(r) + g(r) \cos p\Theta + h(\Theta)$$

where h is a function of Θ only. Differentiating with respect to Θ ,

$$\mu_0 \left(\frac{\partial M}{\partial \Theta} \right) = -pg(r) \sin p\Theta + dh/d\Theta$$

so that
$$\mu_0 \left(\frac{\partial M}{\partial \Theta} \right)_{r=R} = -pg(R) \sin p\Theta + dh/d\Theta = 0$$

whence $dh/d\Theta = pg(R) \sin p\Theta$, and integrating with respect to Θ , $h = -g(R) \cos p\Theta + \alpha$, where α is a constant. Hence

$$\mu_0 M = [f(r) + \alpha] + [g(r) - g(R)] \cos p\Theta$$

since

$$M_{r=R} = 0, \quad \alpha = -f(R)$$

and
$$\mu_0 M = [f(r) - f(R)] + [g(r) - g(R)] \cos p\Theta$$

Writing $F(r)$ for $[f(r) - f(R)]$ and $G(r)$ for $[g(r) - g(R)]$

$$\mu_0 M = F(r) + G(r) \cos p\Theta$$

where $F(R)$ and $G(R)$ are both zero.

Differentiating,

$$\mu_0 \left(\frac{\partial M}{\partial \Theta} \right) = -pG(r) \sin p\Theta, \quad \mu_0 \left(\frac{\partial^2 M}{\partial \Theta^2} \right) = -p^2 G(r) \cos p\Theta$$

$$\mu_0 \left(\frac{\partial M}{\partial r} \right) = \frac{dF}{dr} + \frac{dG}{dr} \cos p\Theta,$$

$$\mu_0 \left(\frac{\partial^2 M}{\partial r^2} \right) = \frac{d^2 F}{dr^2} + \frac{d^2 G}{dr^2} \cos p\Theta = 0$$

Substituting in eqn. (11)

$$\begin{aligned} \frac{d^2 F}{dr^2} + \frac{d^2 G}{dr^2} \cos p\Theta + \frac{1}{r} \frac{dF}{dr} \\ + \frac{1}{r} \frac{dG}{dr} \cos p\Theta - \left(\frac{p}{r} \right)^2 G(r) \cos p\Theta = 0 \end{aligned}$$

for all values of Θ and of r between 0 and R .

In particular, when $p\Theta = \pi/2$ so that $\cos p\Theta = 0$,

$$\frac{d^2 F}{dr^2} + \frac{1}{r} \frac{dF}{dr} = 0$$

Writing F' for dF/dr , this becomes $dF'/dr = -F'/r$ or $dF'/F' = -dr/r$, and integrating,

$$\log_e F' = -\log_e r + \log_e \beta$$

where $\log_e \beta$ is a constant.

Hence $dF/dr = F' = \beta/r$ and $F = \beta \log_e r + \gamma$, where γ is a constant. Since $F(R) = 0$, $\gamma = -\beta \log_e R$ and $F = \beta \log_e (r/R)$. Now $dF/dr = df/dr = (1/R)\beta(R/r) = \beta/r$ so that $(df/dr)_{r=R} = -B_m = \beta/R$, whence $\beta = -RB_m$ and

$$F(r) = -RB_m \log_e (r/R) = RB_m \log_e (R/r)$$

Again, since $\frac{d^2F}{dr^2} + \frac{1}{r} \frac{dF}{dr} = 0$,

$$\frac{d^2G}{dr^2} \cos p\Theta + \frac{1}{r} \frac{dG}{dr} \cos p\Theta - \left(\frac{p}{r}\right)^2 G(r) \cos p\Theta = 0$$

for all values of r and Θ ; hence

$$\frac{d^2G}{dr^2} + \frac{1}{r} \frac{dG}{dr} - \left(\frac{p}{r}\right)^2 G(r) = 0$$

This is a homogeneous linear equation of the second order

which can be solved by putting $G = Pr^n$, when the equation becomes

$$n(n-1)Pr^{n-2} + nPr^{n-2} - p^2Pr^{n-2} = 0$$

whence $n^2 - p^2 = 0$, so that $n = \pm p$ and $G = Pr^p + QR^{-p}$.

Since $G(R) = 0$, $PR^p + QR^{-p} = 0$ or $Q = -pR^{2p}$. Differentiating, $dG/dr = p(Pr^{p-1} - QR^{-p-1})$ and $(dG/dr)_{r=R} = (dg/dr)_{r=R} = p(PR^{p-1} - QR^{-p-1}) = B_c$. Hence, substituting for Q , $2pPR^{p-1} = B_c$, so that $P = B_c/2pR^{p-1}$ and $Q = -B_c/2pR^{-p-1}$.

$$\begin{aligned} \text{Thus } G(r) &= (RB_c/2p)[(r/R)^p - (r/R)^{-p}] \\ &= - (RB_c/2p)[(R/r)^p - (R/r)^{-p}] \end{aligned}$$

and

$$\mu_0 M = RB_m \log_e (R/r) - (RB_c/2p)[(R/r)^p - (R/r)^{-p}] \cos p\Theta$$

Let $(R/r)^p = \varepsilon^\phi$ so that $\phi = \log_e (R/r)^p$ and $\log_e (R/r) = \phi/p$, then

$$\begin{aligned} \mu_0 M &= (RB_m/p)\phi - (RB_c/p)\frac{1}{2}(\varepsilon^\phi - \varepsilon^{-\phi}) \cos p\Theta \\ &= (RB_m/p)\phi - (RB_c/p) \sinh \phi \cos p\Theta \end{aligned}$$

or, alternatively,

$$\mu_0 M = (RB_m/p) \log_e (R/r)^p - (RB_c/p) \sinh [\log_e (R/r)^p] \cos p\Theta \quad (13)$$

PROPERTIES OF SYNTHETIC QUARTZ OSCILLATOR CRYSTALS

By C. S. BROWN, B.Sc., A.Inst.P., and L. A. THOMAS, B.Sc., F.Inst.P.

(The paper was first received 13th April, and in revised form 28th July, 1956. It was published as an INSTITUTION MONOGRAPH in October, 1956.)

SUMMARY

A process for the growing of quartz crystals artificially is briefly described. This so-called synthetic quartz, which is being grown reproducibly as well-formed crystals weighing about 135 grammes, possesses electrical and mechanical properties which closely follow those of the natural Brazilian quartz commonly used for piezo-electric purposes. The high overall quality of synthetic quartz, its freedom from twinning and other faults and its regular shape and size make it a convenient material to handle with the minimum of wastage.

Measurements have been carried out to compare the properties of oscillator crystals cut from natural and synthetic quartz. The behaviour of the AT- and BT-cuts has been investigated thoroughly, while some results have also been obtained on units of the CT-, DT- and 5° X-cuts. Most of the parameters measured show no appreciable difference between the values for natural and synthetic quartz. Two notable exceptions are a small difference in the frequency/temperature characteristics of AT-cut units, whether operated at fundamental or at overtone frequencies, and an apparently reduced activity of face-shear-mode units of the CT- and DT-cuts from the synthetic material.

The significance of these results is discussed, together with the possibility of more accurate measurements to compare the intrinsic mechanical losses in natural and synthetic quartz.

LIST OF SYMBOLS

- R_m, L_m, C_m = Resistance, inductance and capacitance in the motional arm of the equivalent network of an oscillator crystal.
 C_0 = Static capacitance of the crystal measured at a frequency remote from a resonance.
 C'_0 = Capacitance added externally to the crystal unit, either in series or in parallel.
 R_s = Equivalent series resistance, approximately equal to R_m .
 R_p = Equivalent parallel resistance.
 f_s = Series resonant frequency.
 f_a = Parallel resonant frequency.
 ω = Angular frequency = $2\pi f$.
 N = Frequency constant.
 f_T = Frequency at temperature T .
 Δ = Deviation of frequency at temperature T from the frequency at temperature T_0 , parts in 10^6 .
 A, B, C, \dots = Constants in the power series $A(T - T_0) + B(T - T_0)^2 + C(T - T_0)^3 + \dots$
 r = Capacitance ratio = C_0/C_m .
 K = Inductance constant.
 T_{max}, T_{min}, T_0 = Characteristic temperatures of the frequency/temperature curve for an AT-cut crystal.
 θ = Orientation, i.e. complement of angle between plate normal and optic axis of crystal.

(1) INTRODUCTION

The high level of production of quartz crystals attained during the Second World War¹ has brought about a substantial peace-

time demand sustained by the almost exclusive use of quartz vibrators for the control and selection of frequencies in most types of telecommunication equipment other than domestic broadcast receivers. Although quartz is one of the more abundant constituents of the earth's surface, large crystals of piezo-electric quality occur in reasonable quantities in only a few isolated regions. In particular, Brazil has hitherto provided almost all the quartz used in Britain and the United States.

The scarcity of high-grade Brazilian quartz during the war prompted investigations aimed at conserving natural quartz, for example, by the removal of undesirable twinning,^{2,3} and at seeking an alternative source either of quartz or of another piezo-electric material. The untwinning process showed promise, but it had practical defects, and in any event it could only be applied to one of the two common types of twinning, which usually co-exist. Consequently, more emphasis has been placed on research for new raw materials.

Workers in the United States took the lead in the investigation of possible alternatives to quartz, and their search covered both natural minerals⁴ and crystals which grow readily from water solutions.⁵ Although some of the synthetic materials were found to be suitable for use as electro-mechanical transducers or in filter networks,^{6,7,8} none could be considered as a suitable replacement for quartz in oscillator applications. In general, the variation of frequency with temperature for these water-soluble materials is at least 10 times that for quartz.

In the meantime, a start was made in the United Kingdom,⁹ and, as was discovered later, in Germany,¹⁰ on the problem of growing large quartz crystals artificially. Improved growing processes met with considerable success both in the United Kingdom¹¹ and in the United States.^{12,13} However, it should be noted that the results reported here refer solely to material grown by the process which was developed in this country and which is briefly described in Section 2. Although the methods adopted by other workers in this field are broadly similar, certain fundamental differences exist which may affect the crystal and the performance of oscillator units made from it.

(2) METHOD OF GROWTH

A historical summary of the attempts to grow quartz has been given elsewhere.¹¹ Mineralogists have been growing small crystals of quartz for over a hundred years, but with the exception of Spezia¹⁴ at the turn of the century, none has grown crystals of greater than microscopic dimensions until recent times.

Owing to the glass-forming properties of quartz and its allotropic modifications at higher temperatures, it is not practicable to grow large crystals from the melt; that is, the flame fusion (Verneuil) technique and the well-known crucible methods are ruled out. Growth from the vapour phase is also excluded on similar grounds. The method adopted therefore depends on growth from solution. Moreover, as quartz is virtually insoluble in aqueous solutions under ordinary ambient conditions, it is necessary to employ high temperature and pressure to achieve sufficient solubility on which to base a workable process. These so-called hydrothermal conditions are closely related to the

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

conditions that exist a few miles beneath the earth's crust, and they may reproduce approximately the environment in which rock crystal has been formed in nature. At temperatures in the region of 400°C and pressures of 1000 atmospheres (about 7 tons/in²) upwards, quartz is soluble in water and considerably more soluble in alkaline solutions such as sodium carbonate. If a solution saturated with quartz under these conditions is cooled slightly, the solution becomes supersaturated and quartz will deposit on any seed suspended in the liquid. This is the basic process used by Spezia, from which the most successful modern methods have evolved.

A diagram of the apparatus used by the authors is shown in Fig. 1. A steel autoclave constructed to withstand high pressures

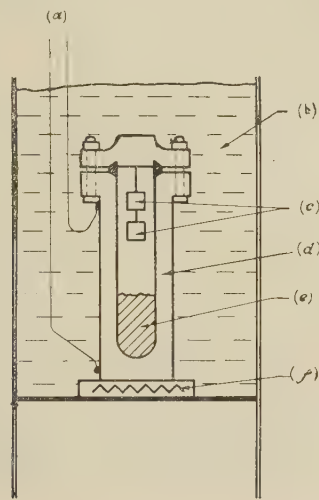


Fig. 1.—Furnace and autoclave for growing synthetic quartz crystals.

- (a) Leads to thermocouples.
- (b) Micaceous thermal insulation.
- (c) Seed crystals.
- (d) Autoclave.
- (e) Crushed melting-grade quartz.
- (f) Hot-plate.

contains seed crystals suspended by fine platinum wires from the lid, and a supply of crushed siliceous mineral to act as the nutrient material at the bottom. The autoclave is about 80% filled with an aqueous solution and is sealed by a mild-steel lens ring placed between conical surfaces on the lid and body. The furnace consists of a hot-plate on which the autoclave stands surrounded by micaceous-flake thermal insulation. In this way a temperature gradient is established along the autoclave so that it is cooler in the upper region, surrounding the seeds, than at the bottom where the nutrient material is located.

When the autoclave is heated the liquid expands to fill the available space completely at a temperature of about 300°C . Any further increase in temperature causes very large pressures to be developed owing to the low compressibility of the liquid, the pressure being a function of the temperature and of the percentage of the space occupied by the liquid at room temperature. The temperature at the base of the autoclave is controlled at about 400°C , and the top is allowed to reach an equilibrium value about 40°C lower. The true temperature drop within the autoclave is certainly much less than this, but no satisfactory method has been found to determine it. Under these conditions quartz is continuously dissolved from the nutrient material and recrystallized on the seed crystal to form single crystals.

In the standard process the seed crystals consist of thin plates of natural or previously grown synthetic quartz cut perpendicular to the optic axis (i.e. Z-cut). The nutrient material is prepared

from small pieces of melting-grade quartz, of the type used in the manufacture of silica glass, and this is employed in conjunction with an aqueous solution containing 10% sodium carbonate by weight. For a given size of autoclave, the rate of growth of the crystal is then dependent primarily on pressure, temperature and temperature differential. The main control of temperature is effected near the base of the autoclave, and it is noteworthy that the simplicity of the furnace arrangement has been achieved without sacrificing the reproducibility of the thermal conditions when the apparatus is dismantled and reassembled. The temperature drop depends on a number of factors, which it is not proposed to discuss here, but they include the effect of the particle-size distribution and quantity of nutrient material in retarding convection and diffusion processes. As implied in the preceding paragraph, the pressure is determined by the initial degree of filling and by the temperatures inside the autoclave. In order to obtain crystals of the highest quality these controlling parameters must be set so that the rate of growth falls below a certain maximum, which is about 0.5 mm per day on each side of the seed, measured in the direction of the optic axis.

Since the linear rate of growth is limited to this maximum it follows that the size of crystal that can be grown is essentially determined by the size of the seed and the duration of the growing process. The latter is decided in practice by the optimum compromise between the need of the application (e.g. the size of oscillator plate to be cut) and the diminishing return arising from the tapering of the crystal as the development of the pyramidal terminations proceeds. It has been found convenient for the majority of applications to employ 35×50 mm seed plates, two of which are readily contained in an autoclave having an internal bore of 63 mm and a capacity of 750 ml. A growth of up to 15 mm then takes place on each side of these seeds in a period of a month, the resulting crystals weighing about 135 g

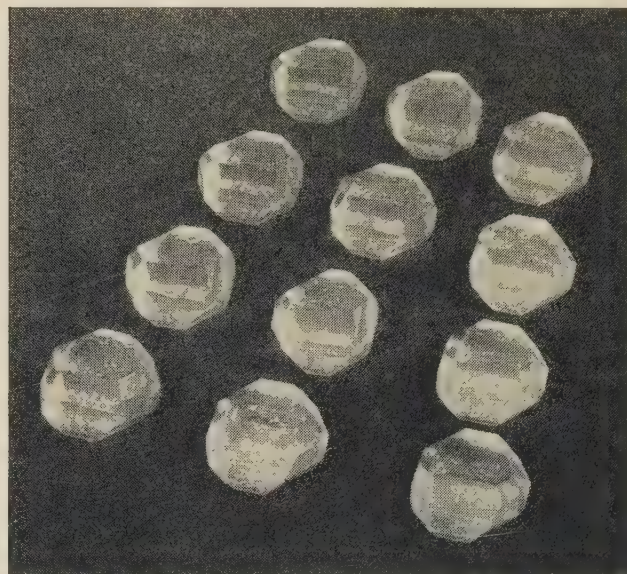


Fig. 2.—Synthetic quartz crystals from pilot plant.

each. Fig. 2 shows a number of these crystals grown in a pilot plant set up for this purpose.

For the economic operation of the pilot plant, the seed crystals are used repeatedly. The crystal is cut into two 'blocks', which possess reference faces perpendicular to the Z and X (diad) axes, by sawing adjacent to the original seed-growth interfaces. The blocks each weigh about 50 g; Fig. 3 shows a reclaimed seed and the blocks associated with it.

It will be seen that the standard process effectively transforms a low-grade quartz, which because of its size and quality is unsuitable for piezo-electric use, into large near-perfect crystals. However, the melting-grade quartz has to be imported, and it was thought desirable to develop other growing techniques using alternative nutrient materials occurring freely in the United Kingdom. In order to employ such indigenous raw materials it is necessary to modify the conditions of growth, including, in many instances, the composition of the aqueous solution. These investigations are still in hand, but satisfactory processes based on the use of flint and quartzite have already been developed. Fig. 4 shows typical crystals grown from these materials in comparison with a crystal grown on a seed of the same size using melting-grade quartz as the nutrient. In the course of this work it has become clear that impurities, especially aluminium, can play a major role in the mechanism of growth.¹⁵

Although Z-cut seeds have normally been used by the authors for both the standard and the modified processes, some experimental work on other seed orientations has been carried out. It is worth while noting that this work may have special importance from the standpoint of the influence of impurities on the physical properties of quartz crystals, because the ease with which impurity atoms are incorporated in the quartz lattice depends among other factors on the orientation of the growing surface.

(3) FABRICATION OF OSCILLATOR CRYSTALS FROM SYNTHETIC QUARTZ

(3.1) Comparison of the Usability of Natural and Synthetic Quartz

The usability of natural quartz, that is, the number of oscillator crystals of a given type that can be cut from a kilogramme of material, is dependent on the size of the oscillator crystal to be cut, the size and shape of the parent crystal, the perfection of the parent crystal, and the cutting techniques used. The trend towards miniaturization in telecommunication equipment has meant that, where possible, the sizes of oscillator crystals have tended to become smaller in the last 10 years. This trend is most noticeable in high-frequency thickness-shear-mode crystals, which represent a large proportion of the units produced in this country. This decrease in the size of oscillator crystals has meant that smaller parent crystals can be used. Whereas the usual weight of parent crystal used here during the war was 1–2 kg, it is now common to use material in the 300–500 g range. This material is commonly irregular, i.e. it is not bounded by natural faces and it may have a wasteful shape. The most important factors affecting the usability, however, are the following imperfections, which are very common in natural quartz:

- (a) Electrical or Dauphiné twinning.
- (b) Optical or Brazil twinning.
- (c) Cracks and veiling.
- (d) Bubbles.
- (e) 'Blue needles.'

These and other defects have been discussed in detail elsewhere.¹⁶ Although most of them can be observed in the parent crystal before cutting, electrical twinning can be detected only after etching in hydrofluoric acid. It is of interest to note that the highest grade of natural quartz commonly available for piezo-electric use is specified to contain not less than 60% usable material, taking no account of electrical twinning.

It is therefore not surprising that, using conventional cutting techniques, the number of 12-mm-diameter discs of less than 1 mm thickness which can be cut from 1 kg of natural quartz is, on average, only about 100.

Synthetic quartz grown by the technique described in Section 2 possesses a number of advantages compared with natural quartz.

When grown on a twin-free seed crystal the deposit is itself free of twinning. Similarly, the gross defects which are common in the natural material are not found in synthetic quartz when it is grown at the optimum rate. Certain minor defects do occur (Section 3.2), but it is found that these have little or no effect on the usability. Since the size of crystal grown is controlled by the size of the seed and by the duration of the growth cycle, the crystal can be made to give the minimum wastage when used to cut a given size of oscillator blank. The blocks of synthetic quartz shown in Fig. 3 are intended primarily for the fabrication

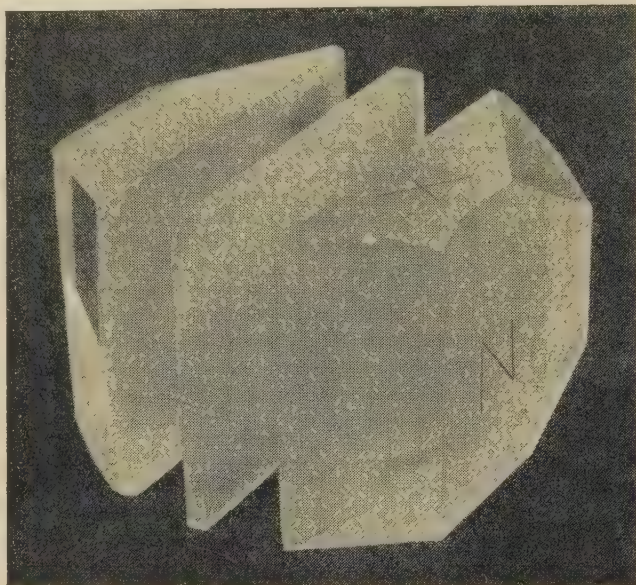


Fig. 3.—Synthetic quartz blocks and reclaimed seed.

of high-frequency thickness-shear-mode crystals of 10–14 mm diameter. Approximately 16 such blanks can be cut by conventional techniques from each block, giving a usability of 320 blanks per kilogramme of block material, or 240 blanks per kilogramme of material as grown, before the seed crystal is recovered. This will be seen to compare favourably with the figure of 100 blanks per kilogramme quoted for natural quartz.

There are other less tangible, but none the less real, advantages in the use of synthetic quartz. The regular shape and size of the synthetic quartz blocks, together with the ease of orientation due to the presence of a number of crystallographic faces, make this material simple to handle in a manner readily capable of mechanization. In addition, this type of block can be handled easily by the new techniques developed for cutting quartz with much reduced wastage. One such device which has been used with considerable success is the annular cutting wheel developed by the Post Office Engineering Department.¹⁷ The yield with this technique can be greater than 60 blanks per block for thickness-shear-mode plates intended for third-overtone operation in the 50 Mc/s region.

(3.2) Imperfections in Synthetic Quartz

It has been pointed out already that the imperfections which are prevalent in natural quartz are not found in synthetic quartz grown under optimum conditions. Certain defects are found which are believed, however, not to affect the overall usability of the material to any great extent.

In nearly all synthetic quartz the most common defects are minute bubbles distributed through the growth. The bubbles are two-phase liquid-vapour inclusions varying in size—mostly in the range 5–50 μ . The concentration of these bubbles varies

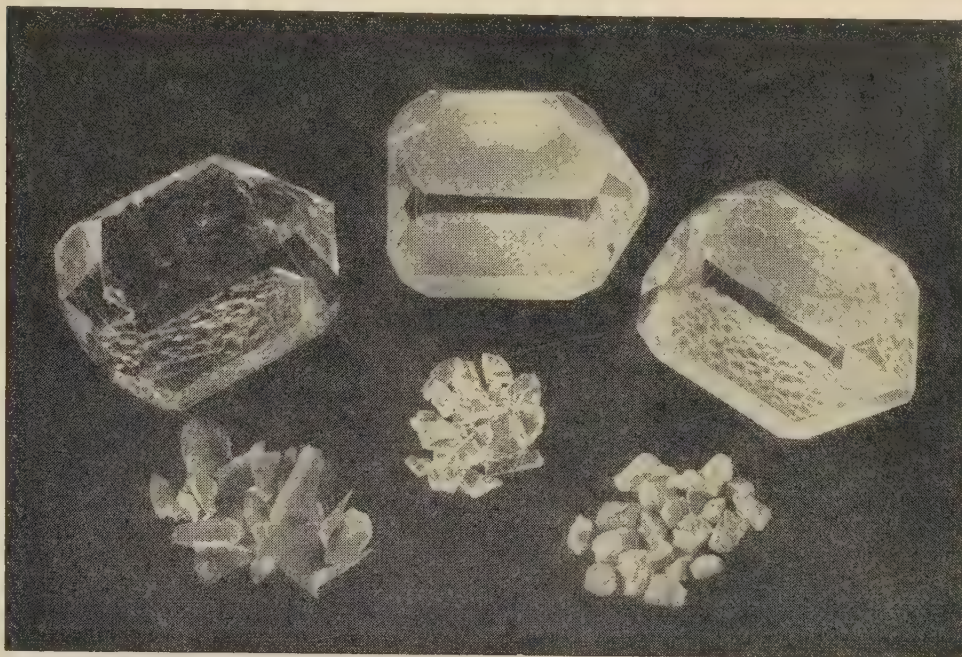


Fig. 4.—Crystals grown from flint (left), melting-grade quartz (centre) and a quartzite (right) with their respective nutrient materials.

considerably from block to block in what appears to be a random manner. Evidence will be given to show that these bubbles, even in relatively large concentrations, do not adversely affect the performance of thickness-shear-mode crystal units.

All the other defects are isolated cracks or needle-like inclusions. It is believed that under optimum conditions these imperfections should not occur, but at present it is not possible to control them completely. It is known that they occur more readily when reclaimed seeds, i.e. seed crystals which have been used previously, are used again. It is probable that better-quality crystals could be obtained if new seeds were used in every growth cycle, and experiments to grow synthetic quartz seeds on an economic basis are being carried out. However, the cracks and inclusions occur either just below the final growth surface or near the edges of the block, so that they have little or no effect on the usability of the material.

(4) PROPERTIES OF OSCILLATOR CRYSTALS

The properties of piezo-electric crystals of any orientation can be derived from a knowledge of the tensor properties (e.g. elastic and piezo-electric coefficients) and their temperature derivatives. This work has been carried out for natural quartz by a number of investigators, including particularly Mason.^{18, 19} Preliminary investigation has shown that, to a first order of approximation, synthetic quartz behaves similarly to the natural crystal. A determination of the individual tensor properties has, therefore, not been made for synthetic quartz, but a detailed study has been carried out on the practical properties of a number of the well-known zero-temperature-coefficient cuts. In this way a few minor differences have been found.

(4.1) Parameters and Methods of Measurement

Van Dyke²⁰ has shown that near resonance a piezo-electric vibrator can be represented by the equivalent network shown in Fig. 5. The precise behaviour of a crystal depends on the type of circuit in which it is to be used, but the following parameters are of importance in determining its performance.

Activity: Activity is a measure of the vibration amplitude of the crystal element in a given mode of vibration under specified

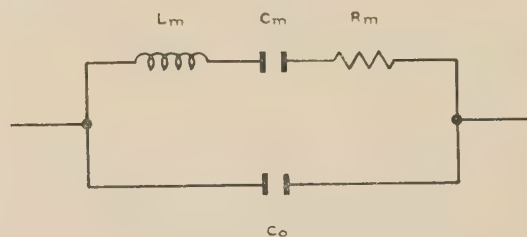


Fig. 5.—Equivalent network of a piezo-electric crystal near an isolated resonance.

conditions. It can be measured in terms of the constants of the equivalent network.

For crystals used in tuned oscillators operating at series resonance, e.g. low-frequency units or overtone-mode thickness-shear units, activity is specified in terms of the equivalent series resistance, R_s , which is approximately equal to R_m . The equivalent series resistance can be measured in a π -network by a substitution method.

Fundamental thickness-shear-mode units, normally in the frequency range 1–20 Mc/s, are commonly operated in aperiodic circuits in the parallel-resonant condition. Activity is then measured in terms of equivalent parallel resistance, R_p . The equivalent parallel resistance has been measured at known input capacitance C_0 in a Type 193 test set.²¹ In terms of the equivalent circuit,

$$R_p = \frac{1}{\omega^2(C_0 + C_0')^2 R_m}$$

Frequency Constant, N : For extensional and shear modes the frequency, f , and frequency-controlling dimension, l , are approximately related by the equation

$$fl = N$$

For a given cut and mode of vibration, N is a constant related to the elastic constants of the material, and is expressed in kilocycle-millimetres per second.

Frequency/Temperature Coefficient: In general, the frequency can be expressed as a power series in terms of temperature thus:

$$\Delta = \frac{f_T - f_0}{f_0} = A(T - T_0) + B(T - T_0)^2 + C(T - T_0)^3 + \dots$$

where f_0 = Frequency at temperature T_0 ;

f_T = Frequency at temperature T ;

and A , B and C are coefficients dependent on orientation and mode of vibration. Δ is expressed in parts in 10^6 per degree Centigrade.

In general, A is large compared with the higher-order coefficients, and hence the frequency approximates to a linear function of temperature. However, there are certain orientations²² for which A is zero, and these are known as zero-temperature-coefficient cuts. For example, there are the AT- and BT-cuts for thickness-shear-mode operation and the CT- and DT-cuts for face-shear-mode operation. For the BT-, CT- and DT-cuts the coefficient B is large, so that the frequency/temperature characteristic is of the quadratic form. Over a small range of orientation T_0 can be considered to vary while A remains zero, so that the turnover temperature (the temperature at which the temperature coefficient is zero) can be varied.

The AT-cut is unique among the above cuts because B is zero over a wide range of orientations on either side of the orientation at which A vanishes. Over the range for which B is zero C can be considered almost constant whereas A varies rapidly. Consequently, by adjusting the orientation, a family of cubic curves can be obtained for which the point of inflection, T_0 , remains approximately constant.

Capacitance Ratio, r : In terms of the equivalent circuit,

$$r = C_0/C_m$$

The ratio of capacitances is a property of the material; for a given cut, mode of vibration and electrode configuration, it is independent of the frequency and of the mounting. The lower the value of r the greater is the piezo-electric coupling between the electrical and mechanical effects.²³

The series and parallel resonant frequencies, f_r and f_a respectively, when the crystal is operated with an added shunt capacitance C'_0 , are given by the relation

$$\frac{f_a - f_r}{f_r} = \frac{C_m}{2(C_0 + C'_0)}$$

The motional capacitance, C_m , can be determined by measurements of $f_a - f_r$ with different values of added shunt capacitance. The capacitance ratio is then obtained by measuring the static capacitance at low frequency.

Inductance, L_m : In the equivalent circuit,

$$L_m = \frac{1}{4\pi^2 f_r^2 C_m}$$

The inductance L_m is dependent on the dimensions and mounting of the crystal as well as on the cut and mode of vibration. For a given mode of vibration the inductance is related to the dimensions by a constant K .

For the extensional mode (fully plated crystal), $L_m = \frac{Klt}{w}$.

For the face-shear mode (fully plated crystal), $L_m = Kt$.

For the thickness-shear mode, $L_m = \frac{Kt^3}{A}$, where A is the electrode area.

Q-factor: In the equivalent circuit,

$$Q = \frac{2\pi f_r L_m}{R_m}$$

The Q-factor is a measure of the losses in the vibrating system. The sources of mechanical dissipation are:

- (a) Internal friction in the body of the crystal.
- (b) Internal friction in the electrode and at the surface of the crystal.
- (c) Radiation of mechanical energy into the mounting system.
- (d) Radiation of mechanical energy into the surrounding medium.

The parameters defined above are sufficient to determine the properties of an oscillator crystal under working conditions. The most important factors which must be controlled are the frequency/temperature coefficient and the crystal activity. The former is critically dependent on crystal orientation, while the activity and Q-factor are chiefly dependent on the quality of the material and on the crystal mounting. It is well known that under normal manufacturing conditions the activity of a given type of oscillator crystal can vary over a considerable range owing to uncontrollable factors. Thus, to compare the activities of certain types of oscillator crystals made from both natural and synthetic quartz, it is essential to make large batches of units which can be subjected to statistical analysis. Moreover, where possible, the units from both materials should be manufactured under identical conditions. It should be noted, however, that high-frequency oscillator crystals are very sensitive to macroscopic defects and, in the development of the synthetic quartz process, many defects have first been observed by their effect on the activity of thickness-shear-mode units fabricated from the material.

(5) RESULTS OF PARAMETER MEASUREMENTS

The results of measurements on batches of AT-, BT-, CT- and DT-cut plates and on 5°X-cut bars are considered in this Section.

(5.1) AT-Cut

The AT-cut is coming into increasing use for fundamental-mode operation in the medium-frequency range (1–20 Mc/s), and is already predominant for overtone operation at higher frequencies. Its cubic frequency/temperature characteristic makes it possible to design crystal units with a very small frequency change over a wide temperature band; for example, ± 50 parts in 10^6 from -55 to $+90^\circ\text{C}$.

(5.1.1) Activity.

Six synthetic quartz blocks chosen as a representative sample have been fabricated under factory conditions to produce about 100 4.95 Mc/s AT-cut units, orientation (YXI) $+35^\circ 15'$ (I.R.E. notation²⁴). Mounted by standard techniques in style-C metal holders,²⁵ the units are found to show the expected wide spread in activities, as is shown by the distribution curve in Fig. 6.

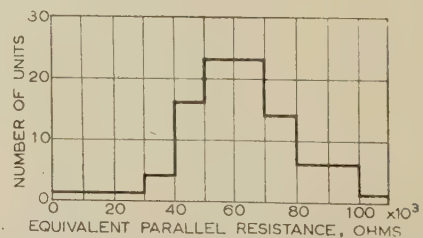


Fig. 6.—Values of equivalent parallel resistance, R_p , for a batch of 4.95 Mc/s AT-cut synthetic quartz crystal units.

The mean effective parallel resistance is 62 kilohms, with a standard deviation of 18 kilohms.

The activity of higher-frequency fundamental- and overtone-mode units is critically dependent on the fabrication and mounting

techniques adopted. 50 Mc/s third-overtone units made from synthetic quartz blanks finished on No. 304 emery and using evaporated gold electrodes have values of R_s in the range 20–30 ohms. Similar figures are obtained for natural quartz units.

(5.1.2) Frequency/Temperature Characteristics.

Measurements have been made on a large number of synthetic quartz units operating on fundamental and third-overtone modes, and preliminary results are available on the fifth-overtone mode. Most of the measurements are for units with a fundamental frequency of about 17 Mc/s with corresponding third- and fifth-overtone frequencies in the 50 Mc/s and 85 Mc/s region. Comparison measurements have also been carried out on natural quartz units operating on third-overtone mode at 50 Mc/s. Bechmann and Hale^{26,27} have recently published results for synthetic quartz grown by a somewhat different technique¹³ employing a seed cut parallel to a minor rhombohedral face (r-cut) in place of a seed perpendicular to the optic axis (Z-cut) as described by the authors in Section 2. Bechmann and Hale's measurements have been carried out on 2 Mc/s bevelled AT-cut discs operated at their fundamental, third- and fifth-overtone frequencies. These results cannot be compared directly with those obtained at higher frequencies, since it is known that, at low frequencies, the form of the frequency/temperature characteristic is frequency-dependent.

As mentioned previously, the frequency/temperature characteristic of an AT-cut unit can be represented approximately by the equation

$$\Delta = A(T - T_0) + C(T - T_0)^3$$

Over a limited range of orientation, C is found to be constant while A is a linear function of the orientation, θ , so that

$$A_0 = (\theta - \theta_0) \frac{\partial A}{\partial \theta}$$

where θ_0 = Orientation for which $A = 0$.

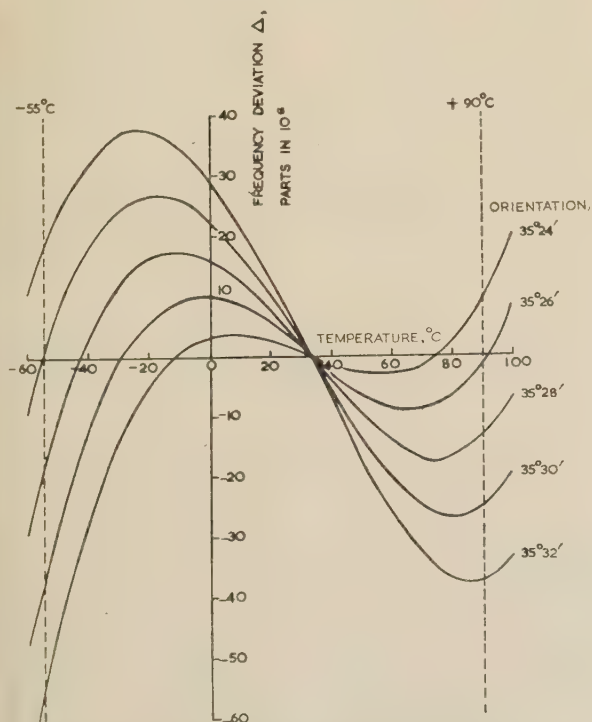


Fig. 7.—Idealized frequency/temperature characteristics of third-overtone 50 Mc/s AT-cut units from synthetic quartz.

T_0 = Temperature of the point of inflection of the cubic curve.

Figs. 7 and 8 show idealized characteristics of third-overtone 50 Mc/s units cut from synthetic and natural quartz respectively, for orientations from $35^\circ 24'$ to $35^\circ 32'$. The differences in the characteristics are more clearly seen in Fig. 9, where the temperatures of the maxima, T_{max} , and minima, T_{min} , and inflection, T_0 , are plotted against orientation for synthetic and natural quartz. From the empirical formula for the frequency deviation

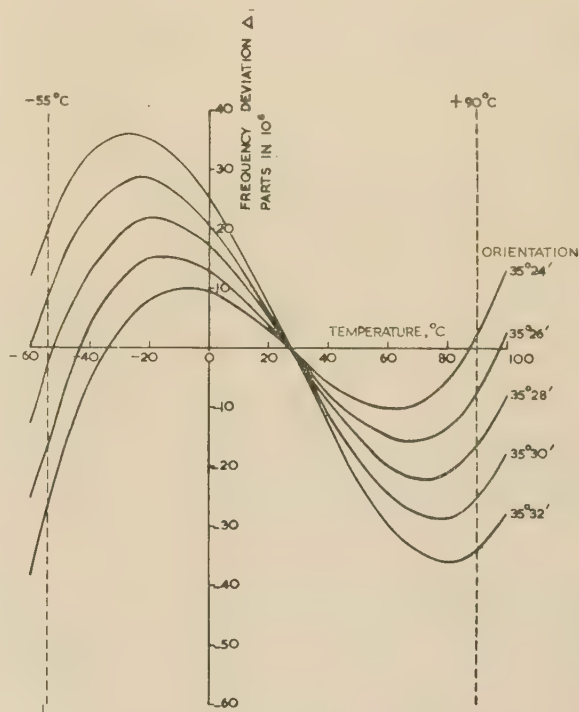


Fig. 8.—Idealized frequency/temperature characteristics of third-overtone 50 Mc/s AT-cut units from natural quartz.

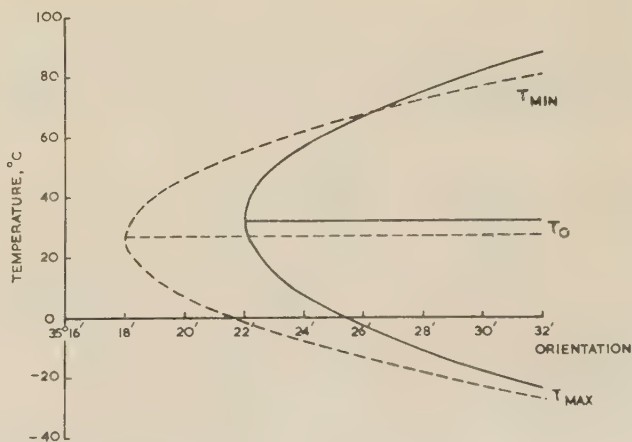


Fig. 9.— T_{max} , T_{min} and T_0 as functions of orientation for third-overtone 50 Mc/s AT-cut crystal units from synthetic and natural quartz.

————— Synthetic quartz,
----- Natural quartz.

tion Δ , the maximum and minimum temperatures T_m are defined by

$$(T_m - T_0)^2 + \frac{1}{3} \frac{1}{C} \frac{\partial A}{\partial \theta} (\theta - \theta_0) = 0$$

That is, T_{min} and T_{max} lie on a parabola with orientation as the abscissa. From Fig. 9 it will be seen that θ_0 , T_0 and the ratio $\frac{1}{C} \frac{\partial A}{\partial \theta}$, which defines the shape of the parabola, all have slightly different values for natural and synthetic quartz. Table 1 gives the values of θ_0 , T_0 , $\frac{\partial A}{\partial \theta}$ and C for natural and synthetic quartz and, for comparison, the results given by Bechmann and Hale on low-frequency units.

The values for the frequency swing are mean values from experimental results and thus allow for the spread in results found with measurements of this type. Values calculated from the idealized curves in Figs. 7 and 8 are rather lower than these and give the minimum value which is theoretically possible.

It will be seen that the optimum orientations for the two materials are very nearly the same. This is due to the fact that over the temperature range -55 to $+90^\circ\text{C}$ the differences in θ_0 and $\partial A/\partial \theta$ nearly balance one another. This would not necessarily be true for a different temperature range. The difference in the overall frequency swing is largely due to the difference in T_0 , the inflection temperature. It will be seen that the minimum frequency swing will occur when the end temperatures of the temperature range are symmetrically disposed

Table 1

FREQUENCY/TEMPERATURE CHARACTERISTICS OF AT-CUT UNITS

Material	Fundamental frequency	Order of mode	θ_0	T_0	$\partial A/\partial \theta$	C
Synthetic quartz grown on a Z-cut seed	17	Fundamental	$35^\circ 14'$	35	per degree -4.6	parts in 10^6 per $(^\circ\text{C})^3$ 110
		Third overtone	$35^\circ 22'$	32	-6.1	110
		Fifth overtone	$35^\circ 21'$	30	-6.5	120
Natural quartz	17	Third overtone	$35^\circ 18'$	27	-4.1	115
Synthetic quartz grown on an r-cut seed (Bechmann and Hale)	2	Fundamental	$35^\circ 16'$	34	-5.2	115
		Third overtone	$35^\circ 24'$	35	-5.2	110
		Fifth overtone	$35^\circ 26'$	36	-5.2	105
Natural quartz (Bechmann and Hale)	2	Fundamental	$35^\circ 10'$	27.5	-5.2	120
		Third overtone	$35^\circ 18'$	30	-5.2	105
		Fifth overtone	$35^\circ 20'$	31.5	-5.2	100

It will be seen that both θ_0 and T_0 are higher for synthetic than for natural quartz for 50 Mc/s third-overtone units. From the results of Bechmann and Hale it appears that the differences for their material are in the same direction but of rather greater magnitude.

The optimum orientation for a minimum frequency change over a temperature band can be found by inspection of the frequency/temperature characteristic curves in Figs. 7 and 8. Table 2 gives the optimum orientation for third-overtone 50 Mc/s

about T_0 . For the range considered, the ideal inflection temperature is 17.5°C . Thus natural quartz would be expected to give a slightly lower frequency swing, since T_0 is nearer the ideal value. In practice, the small difference found is hardly significant, considering the spread in results for a number of units cut to the same orientation.

(5.1.3) Design Parameters.

The parameters N , r , K and Q defined in Section 4.1 have been measured, and Table 3 gives representative values compared

Table 2

OPTIMUM ORIENTATION FOR 50 Mc/s AT-CUT UNITS OVER THE TEMPERATURE RANGE -55 TO $+90^\circ\text{C}$

Material	Orientation	Overall frequency swing
		Parts in 10^6
Synthetic quartz	$35^\circ 28'$	42
Natural quartz	$35^\circ 27'$	37

units for the temperature range -55 to $+90^\circ\text{C}$, together with values of the overall frequency swing to be expected. Similarly, the optimum orientation for fundamental- or fifth-overtone-mode operation over any temperature can be calculated from the constants in Table 1.

Table 3

DESIGN PARAMETERS FOR THE AT-CUT

Material and frequency	N	r	K	Q
	kc-mm/s		H/cm	
Synthetic quartz:				
4.95 Mc/s	1 660	240	400	100 000
17 Mc/s	1 660	240	400	55 000
Natural quartz:				
17 Mc/s	1 660	230	400	65 000

with those for natural quartz. It must be noted that the parameters are only approximately constant and the purpose of quoting the figures is to show their order of magnitude.

(5.2) BT-Cut

The BT-cut, although rapidly being superseded for many purposes by the AT-cut, still finds use where the frequency tolerance is less severe or for high-stability use when temperature-controlled ovens can be used. In general, a BT-cut unit is easier to make, since, because of the reduced piezo-electric coupling, the units are less susceptible to discontinuities in the activity/temperature characteristic due to coupling with unwanted modes of vibration or with the mounting.

(5.2.1) Activity.

A number of synthetic quartz blocks have been fabricated to make BT-cut crystals in the range 6–10 Mc/s. To determine the effect of some of the minor imperfections discussed in Section 3.2, six blocks have been cut to make 9.68 Mc/s BT-cut units of orientation (YXI) $-49^\circ 20'$ mounted in style-C metal holders. The blocks were chosen to cover a wide range of quality as shown in Table 4, which gives the results of measurements of R_p at 30 μ F input capacitance.

Table 4

EFFECT OF BLOCK QUALITY ON THE ACTIVITY OF
9.68 Mc/s BT-CUT UNITS

Block quality		Number of units	Equivalent parallel resistance (30 μ F input cap.)	
Cracking	Bubbles		Mean	Range
0	1	16	k Ω 18.4	k Ω 14–24.5
0	1	16	14.7	9–18.5
2	2	16	18.5	12.5–25
1	4	15	18.9	13.5–23
0	1	13	18.4	12–21
0	1	15	15.6	8–20.5

Note.—The defects are numbered 0–4 in increasing order of severity.

It will be seen that the defects have no significant effect on the activity and that the units all meet the RCS pass limit for this type, namely 8 kilohms. Similar batches of units made from natural quartz have mean values of R_p ranging from 15 to 22 kilohms.

(5.2.2) Frequency/Temperature Characteristics.

The frequency/temperature characteristic of a BT-cut crystal can be represented by the equation

$$\Delta = B(T - T_0)^2$$

where both B and T_0 are functions of orientation. T_0 , the temperature of the maximum, which is often called the 'turnover temperature', is normally adjusted to be at the mid-point of a wide temperature band or at the oven temperature when used in conjunction with thermostatic control.

Six units have been made from synthetic quartz at each of five orientations from $-49^\circ 40'$ to $-47^\circ 30'$. Fig. 10 shows the variation of B and T_0 against orientation. Mean values of T_0 are plotted, together with the spread over the six units. Measurements at two orientations at the ends of the range using natural quartz give results in good agreement, but show at least twice the spread in values compared with the synthetic crystals over the same small sample.

The BT-cut finds only limited application when operated on an overtone mode owing to its relatively low activity and large frequency drift over an extended temperature range. Measurements of frequency/temperature coefficient have been made on third-overtone-mode BT-cut units at 30 Mc/s. Slightly lower

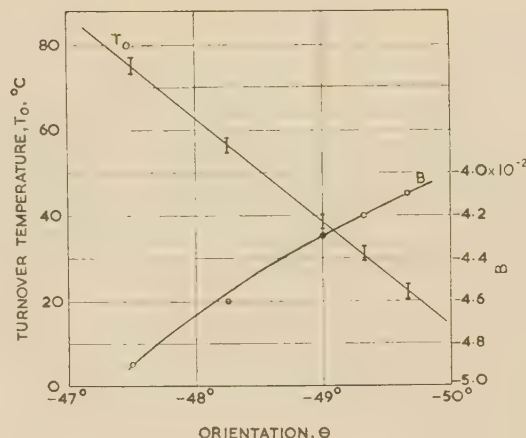


Fig. 10.—Constants B and T_0 as functions of orientation for synthetic quartz BT-cut crystal units.

I indicates spread in results over six units.

values of B and values of T_0 about 20°C lower than for the same crystal at fundamental mode are obtained. This behaviour is in agreement with results on natural quartz units.

(5.2.3) Design Parameters.

Parameters for typical BT-cut units at 10 Mc/s are given in Table 5.

Table 5

DESIGN PARAMETERS FOR BT-CUT UNITS

	N	r	K	Q
	kc-mm/s		H/cm	
Synthetic quartz ..	2 550	710	600	145 000
Natural quartz ..	2 550	680	600	145 000

(5.3) CT- and DT-Cuts

CT- and DT-cut crystals operated on a face-shear mode find considerable use in the range 100–500 kc/s, the DT-cut mostly at the lower end of the range owing to its smaller frequency constant.

(5.3.1) Activity.

Two synthetic quartz blocks chosen at random were fabricated into 308 kc/s CT-units (YXI $+37^\circ 27'$) and 142 kc/s DT-units (YXI $51^\circ 53'$) respectively. Table 6 shows the results of

Table 6

ACTIVITY OF CT- AND DT-CUT UNITS

Type	Synthetic quartz			Natural quartz		
	Number of units	Equivalent series resistance		Number of units	Equivalent series resistance	
		Mean	Standard deviation		Mean	Standard deviation
308 kc/s CT-cut	15	ohms 169	ohms 10.5	90	ohms 89	ohms 17
142 kc/s DT-cut	9	250	17	15	150	18.5

measurements of R_s in vacuo on these units, together with comparison figures for similar units from natural quartz.

The equivalent series resistances are sufficiently consistent, as shown by the standard deviations, to justify the statement that the difference between the results for natural and synthetic quartz is significant. Further tests have been carried out to determine the effect of the concentration of bubbles on the activity of synthetic-quartz CT-cut units. Two blocks, one having a low bubble density and the other an abnormally high concentration, have been fabricated into 308 kc/s CT-cut units. The results are given in Table 7.

Table 7

THE EFFECT OF BUBBLE DENSITY IN SYNTHETIC QUARTZ ON THE ACTIVITY OF CT-CUT RESONATORS

Low bubble density			High bubble density		
Number of units	Equivalent series resistance		Number of units	Equivalent series resistance	
	Mean	Standard deviation		Mean	Standard deviation
18	ohms 129	ohms 11	20	ohms 115	ohms 9.5

The mean equivalent series resistances for both batches in this second test are considerably lower than those found in the original test. Statistical analysis still shows, however, that the difference between natural and synthetic units is significant. Considering the effect of bubble density, the results indicate that the poorer activity of the synthetic-quartz units is probably not associated with the presence of bubbles in the material. It will be seen that any difference in the equivalent series resistances of units made from synthetic quartz with high and low bubble densities is, if anything, in the direction opposite to that which would be expected. It is not yet possible to explain the poorer activities of synthetic quartz operated on a face-shear mode. It should be noted, however, that the difference is of such a magnitude that it should not prejudice the use of synthetic quartz except in applications where the highest possible Q-factor is required.

(5.3.2) Frequency/Temperature Characteristics.

The frequency/temperature characteristic of both the CT- and DT-cut units is parabolic, following the law

$$\Delta = B(T - T_0)^2$$

Results of measurements on natural and synthetic quartz units cut to the same orientations are given in Table 8. The results are seen to be in good agreement.

Table 8

FREQUENCY/TEMPERATURE CHARACTERISTICS OF CT- AND DT-CUT UNITS

Cut	Synthetic quartz		Natural quartz	
	B	T ₀	B	T ₀
CT (+37° 27')	-4.9×10^{-2}	°C 26.5	-5.0×10^{-2}	°C 26.5
DT (-51° 35')	-2.45×10^{-2}	26	-2.5×10^{-2}	27

(5.3.3) Design Parameters.

Parameters for CT- and DT-cut units are given in Table 9.

Table 9

DESIGN PARAMETERS FOR CT- AND DT-CUT RESONATORS

	N	r	K	Q
	kc-mm/s		H/cm	
CT-cut:				
Synthetic quartz ..	3 080	380	230	190 000
Natural quartz ..	3 080	350	225	250 000
DT-cut:				
Synthetic quartz ..	2 060	350	500	130 000
Natural quartz ..	2 060	400	535	220 000

(5.4) +5°X-Cut Bars

In consequence of the results on face-shear-mode resonators, the behaviour of other low-frequency units is being examined. In particular, a small number of 100 kc/s +5°X-cut bars have been made from natural and synthetic quartz.

(5.4.1) Activity.

It was not possible to obtain the same level of reproducibility with bars as was achieved with the face-shear resonators. However, within the limits of accuracy of the measurements, the results given in Table 10 show no significant difference in activity between synthetic and natural quartz.

Table 10

ACTIVITY OF 100 KC/S +5°X-CUT BARS

	Number of units	Equivalent series resistance	
		Mean in vacuo	Range
Synthetic quartz	6	ohms 660	ohms 550-720
Natural quartz	9	640	400-950

(5.4.2) Frequency/Temperature Characteristic.

For values of the ratio of width to length below 0.2, the frequency/temperature characteristic follows the equation

$$\Delta = B(T - T_0)^2$$

A comparison of the constants of this equation for the 100 kc/s bars is given in Table 11.

Table 11

FREQUENCY/TEMPERATURE CHARACTERISTIC OF 100 KC/S +5°X-CUT BARS

	B	T ₀
Synthetic quartz	-3.7×10^{-2}	°C 58
Natural quartz	-3.8×10^{-2}	58

(5.4.3) Design Parameters.

Design parameters for the +5°X-cut bar are also a function of the width-to-length ratio. The parameters are given in Table 12 for a ratio of approximately 0.16.

Table 12

DESIGN PARAMETERS OF $+5^\circ\text{X}$ -CUT BARS ($w/l = 0.16$)

	N	r	K	Q (typical)
	kc-mm/s		H/cm	
Synthetic quartz ..	2810	120	88	130 000
Natural quartz ..	2810	125	88	130 000

(6) DISCUSSION

The results show that synthetic quartz can be used to replace natural quartz in most, if not all, applications; moreover, because of its reproducible shape and size and its freedom from twinning it possesses advantages over the natural mineral. For the majority of the cuts examined, natural quartz can be replaced directly by synthetic quartz, since there is no measurable difference in the properties of the two materials. In the case of the AT-cut, the orientation may have to be modified slightly to obtain the optimum frequency/temperature characteristic for a given operating temperature range. The slightly increased overall frequency-swing for the commonly used temperature range of -55 to $+90^\circ\text{C}$ is likely to be counteracted in practice by the greater consistency in the behaviour of synthetic quartz units cut to the same orientation. This improved reproducibility could be an important advantage for close-tolerance units. In general, Q-factors for natural and synthetic quartz are comparable. The somewhat lower Q-factor found for the synthetic quartz face-shear-mode units is probably unimportant except where they are to be used for applications requiring the highest possible value.

It has been recognized that measurements of Q-factor on practical oscillator units are often insensitive as a guide to the true internal damping in the vibrating crystal. The highest Q-factors found for the types of oscillator unit considered are of the order of 2×10^5 . It is known, however, that Q-factors as high as 5×10^6 are possible with natural quartz. As shown in Section 4.1, the Q-factor of a piezo-electric crystal unit is dependent not only on the internal friction in the body of the material but also on other factors concerned with its use in practice. For example, most of the types in common use have their Q-factors depressed by coupling to the mounting system. In order to obtain a measure of the damping due to the crystal alone, a design must be used in which external effects are reduced to a minimum. Three types of crystal resonator fulfilling this requirement are mentioned below.

(a) *The Essen Z-cut ring.*²⁸ In the form of a 100 kc/s ring this device requires the use of a very large crystal (about $70 \times 70 \times 15$ mm). Crystals of the required size, weighing up to nearly a kilogramme, are now being grown for this purpose. It is not, however, possible to test the standard size of crystal by this method.

(b) *GT-cut crystals.*²⁹ This also requires large crystals.

(c) *Fifth-overtone 5 Mc/s AT-cut units.*³⁰ Using a design described by Warner it is possible to obtain crystals with a Q-factor as high as 2×10^6 with a type of unit which can be cut from the standard synthetic quartz block.

The last-named offers the simplest solution, and an investigation has been started using such crystals as a means of assessing the true internal damping of synthetic quartz.

The difference between the frequency/temperature characteristics of various AT-cut units is another feature of considerable interest. Chi, Hammond and Gerber³¹ have shown that, by the deliberate addition of impurities into the quartz lattice, the

frequency/temperature characteristics of AT-cut units can be considerably modified. In particular, a crystal believed to contain 100 parts of aluminium and 50 parts of germanium in 10^6 , plus other minor impurities, was found to give $T_0 = 75^\circ\text{C}$ and $\theta_0 = +35^\circ 55'$ for fifth-overtone crystals at 29 Mc/s. By comparison with the values quoted in Table 1, it will be seen that the effect of these added impurities has been to shift the values of T_0 and θ_0 further away from those for natural quartz. Considerable evidence has been obtained from spectrographic analyses and the effect of ionizing radiation on quartz³² to show that synthetic quartz as grown by the authors on a Z-cut seed has lower concentrations of impurities such as aluminium than either natural or synthetic quartz grown on an r-cut seed. For example, Table 13 shows the results of spectrographic analyses on typical specimens of natural and synthetic quartz.

Table 13

SPECTROGRAPHICALLY DETERMINED ALUMINIUM IN QUARTZ

Material	Al concentration
	Parts in 10^6
Natural quartz (Brazil)	50-100
Synthetic quartz (Z-cut seed)	<20
Synthetic quartz (r-cut seed)	100-200
Synthetic quartz (doped) ³¹	100*

* Plus 50 parts Ge in 10^6 and other impurities.

The effects on the frequency/temperature characteristic of the AT-cut cannot be explained solely in terms of impurity because the difference between synthetic quartz grown on Z- and r-cut seeds respectively is very much less than the effect of the impurities in the 'doped' quartz measured by Chi and his co-workers. In the latter instance it is possible that one of the impurities other than aluminium has a substantial effect on the properties of the material. It appears, however, to be more likely that the quartz is strained owing to the presence of non-uniformly distributed impurities and that this affects the frequency/temperature characteristic. There is reason to believe that the material grown on a Z-cut seed is largely free of strain.

Under closely controlled conditions it may be possible to introduce appropriate impurities into synthetic quartz, in a regular way, so that the resultant oscillator units give an improved performance compared with natural quartz. There is reason to believe that few elements can be introduced readily into quartz grown on a Z-cut seed. In addition, those elements, such as aluminium, which can be deliberately incorporated in such a crystal tend to be distributed in a non-uniform manner, thus setting up considerable strain which renders the material practically useless for piezo-electric applications. It appears likely that any attempt to study the effects of impurities on the properties of quartz will have to be carried out on material grown on r-cut seeds.

(7) REFERENCES

- (1) BOOTH, C. F., and JOHNS, J. P.: 'The Development of Quartz Crystal Production', *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 899.
- (2) WOOSTER, W. A., WOOSTER, NORA, RYCROFT, J. L., and THOMAS, L. A.: 'The Control and Elimination of Electrical (Dauphiné) twinning in Quartz', *ibid.*, 1947, **94**, Part IIIA, p. 927.
- (3) THOMAS, L. A., and WOOSTER, W. A.: 'Piezocrescence—the Growth of Dauphiné Twinning in Quartz under Stress', *Proceedings of the Royal Society, A*, 1951, **208**, p. 43.

- (4) BOND, W. L.: 'A Mineral Survey for Piezoelectric Materials', *Bell System Technical Journal*, 1943, **32**, p. 145.
- (5) EGLI, P. H.: 'A Survey of Inorganic Piezoelectric Materials', *American Mineralogist*, 1948, **33**, p. 622.
- (6) MASON, W. P.: 'New Low-Coefficient Synthetic Piezoelectric Crystals for use in Filters and Oscillators', *Proceedings of The Institute of Radio Engineers*, 1947, **35**, p. 1005.
- (7) BIRCH, J., FRITH, A. G., FERGUSON, A. C. L., MILES, R. H. A., and WERNER, J. F.: 'The Use of Ethylene Diamine Tartrate for Piezo-electric Filter Elements', *Proceedings I.E.E.*, Monograph No. 182 R, June, 1956 (**103 C**, p. 420).
- (8) MASON, W. P.: 'Piezoelectric Crystals and their Application to Ultrasonics' (Van Nostrand, 1950), pp. 114-233.
- (9) WOOSTER, NORA, and WOOSTER, W. A.: 'Preparation of Synthetic Quartz', *Nature*, 1946, **157**, p. 297.
- (10) NACKEN, R.: 'Interrogation of German Scientists regarding Quartz Crystals and other Piezoelectric Materials', F.I.A.T. Final Report, 1945, No. 641, p. 11.
- (11) BROWN, C. S., KELL, R. C., THOMAS, L. A., WOOSTER, NORA, and WOOSTER, W. A.: 'The Growth and Properties of Large Crystals of Synthetic Quartz', *Mineralogical Magazine*, 1952, **29**, p. 858.
- (12) WALKER, A. C., and BUEHLER, E.: 'Hydrothermal Growth of Quartz Crystals as related to Phase Considerations', *Industrial and Engineering Chemistry*, 1954, **46**, p. 1670.
- (13) HALE, D. R.: 'The Properties of Synthetic Quartz and their Growing Technique', *Brush Strokes*, December, 1952.
- (14) SPEZIA, G.: 'Sull'accrescimento del quarzo', *Atti della R. Accademia delle scienze*, Torino, 1908, **44**, p. 95.
- (15) BROWN, C. S., KELL, R. C., MIDDLETON, P., and THOMAS, L. A.: 'Influence of Impurities on the Growth of Quartz Crystals from Flint and Quartzite', *Nature*, 1955, **175**, p. 602.
- (16) WILLARD, G. W.: 'Raw Quartz. Its Imperfections and Inspection', *Bell System Technical Journal*, 1943, **22**, p. 338.
- (17) THWAITES, J. E., and SAYERS, C. F.: 'A Novel Type of Saw for the Economical Cutting of Quartz Crystals or other Materials', *Post Office Electrical Engineers' Journal*, 1955, **47**, Part 4, p. 1.
- (18) MASON, W. P.: 'Quartz Crystal Applications', *Bell System Technical Journal*, 1943, **32**, p. 178.
- (19) MASON, W. P.: 'Piezoelectric Crystals and their Applications to Ultrasonics' (Van Nostrand, 1950), Chapter 5.
- (20) VAN DYKE, K. S.: 'The Piezoelectric Resonator and its Equivalent Network', *Proceedings of the Institute of Radio Engineers*, 1928, **16**, p. 742.
- (21) BIGGS, A. J., and WELLS, G. M.: 'The Measurement of the Activity of Quartz Oscillator Crystals', *Journal I.E.E.*, 1946, **93**, Part III, p. 29.
- (22) MASON, W. P.: 'Low-Temperature-Coefficient Quartz Crystals', *Bell System Technical Journal*, 1940, **29**, p. 74.
- (23) CADY, W. G.: 'Piezoelectricity' (McGraw-Hill, 1946), Chapter 14.
- (24) 'Standards on Piezoelectric Crystals', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 1378.
- (25) 'Joint Service Standards for Radio Components. List of Standard Crystal Units, Quartz (Oscillators), RCL.271', 1952.
- (26) BECHMANN, R., and HALE, D. R.: 'Electronic Grade Synthetic Quartz', *Brush Strokes*, 1955, **4**, p. 1.
- (27) BECHMANN, R.: 'Frequency-Temperature Behaviour of Piezoelectric Resonators made of Natural and Synthetic Quartz', *Convention Record of the Institute of Radio Engineers*, 1955, **3**, A.9, p. 56.
- (28) ESSEN, L.: 'New Form of Frequency and Time Standard', *Proceedings of the Physical Society*, 1938, **50**, p. 413.
- (29) MASON, W. P.: 'A New Crystal Plate designated the "GT"', *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 220.
- (30) WARNER, A. W.: 'High-Frequency Crystal Units for Primary Frequency Standards', *ibid.*, 1952, **40**, p. 1030.
- (31) CHI, A. R., HAMMOND, D. L., and GERBER, E. A.: 'Effect of Impurities on the Resonator Properties of Quartz', *ibid.*, 1955, **43**, p. 1137.
- (32) BROWN, C. S., and THOMAS, L. A.: 'Response of Synthetic Quartz to X-ray Irradiation', *Nature*, 1952, **169**, p. 35.

THE EFFECT OF MAGNETIC SATURATION ON THE D.C. DYNAMIC BRAKING CHARACTERISTICS OF A.C. MOTORS

By O. I. BUTLER, M.Sc., Member.

(The paper was first received 4th June, and in revised form 1st August, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

It is shown that the effect of magnetic saturation on the d.c. dynamic braking characteristics of a.c. motors can be determined by a mathematical method of analysis which uses the net value of the equivalent alternating magnetizing current, rather than the fractional speed, as the independent variable of functions expressing the values of the braking torque and secondary-circuit current. The method is applicable to both asynchronous and synchronous polyphase machines, and to asynchronous single-phase motors. It is noted that the deep-bar type of squirrel-cage induction motor, in which the rotor resistance varies appreciably with the rotor speed, does not present any special difficulty in the application of the method.

A quantitative relationship is derived for the maximum-torque conditions. The variation of the maximum torque, and of the critical-speed/secondary-resistance ratio, with the direct primary current is determined for values of the latter up to four times the rated primary current, and with zero, normal and excessive values of secondary reactance. It is shown that even when the existence of the normal reactance only is neglected, the maximum torque may be in error to the extent of 16%. Still greater errors arise from the neglect of magnetic saturation, even with the direct primary current limited to the rated value.

It is shown, also, that the torque/resistance characteristics may be derived directly from the torque/speed characteristics.

(1) LIST OF PRINCIPAL SYMBOLS

- V_0 = Induced e.m.f. per secondary phase at synchronous speed, volts.
- I_d = Direct primary current, amp.
- I_1 = Equivalent alternating primary current per phase, referred to secondary circuit, amp.
- I_m = Net magnetizing component of I_1 , amp.
- I_2 = Secondary current per phase, amp.
- m = Number of phases of secondary winding.
- N = Rotor speed, r.p.m.
- N_s = Synchronous speed, r.p.m.
- R_2 = Total secondary-circuit resistance per phase, ohms.
- S = Fractional speed, $= N/N_s$.
- T = Torque, 'synchronous watts'.
- X_m = Magnetizing reactance, V_0/I_m , ohms.
- X_2 = Total secondary-circuit reactance per phase, ohms.

(2) INTRODUCTION

The braking of induction motors by disconnecting the a.c. supply and applying direct current to the primary winding was described as early as 1910 by Rosenberg and Peck,¹ and Helmund.² The method assumed particular significance in 1935 in its application to run-out-table drives for high-speed steel-strip mills,³ and in 1936 in connection with mine-winders.⁴ It is now being used increasingly for other types of drive, such as centrifuges, machine tools and lifts.

In general, the ratio of the direct braking current to the rated alternating motoring current depends upon the type of load. For example, the ratio is of the order of 1.0 to 2.0 in the case of a deadweight load which decelerates the drive immediately the a.c. supply is disconnected; 2.0 to 2.5 in drives where the moment of inertia of the load is negligible compared with that of the rotor; 3.0 to 4.0 in loads possessing appreciable flywheel effect; and still greater values may be encountered in loads which tend to drive the motor beyond its synchronous speed.

It follows that, since the armature-reaction effect of the secondary circuit is reduced as the motor speed falls during braking and is zero at standstill, the magnetic circuit can attain excessive saturation values. In consequence, the predetermination of the braking torque/speed characteristic is seriously in error when saturation is ignored, thereby affecting the correct choice of motor frame size and the precision of the control design.

The increasing application of the method, and the practical interest in the predetermination of the performance characteristics, is reflected in the number of contributions to the solution of the problem of saturation effects published in the last few years.⁵⁻⁹ Latterly, Stul'nikov¹⁰ has shown that the maximum torque conditions can be determined, when the secondary-circuit reactance is neglected, in a manner similar to that put forward by Weissheimer¹¹ and Friedlander,¹² whilst Harrison¹³ has given an accurate graphical method of predicting the torque/speed and torque/resistance characteristics, within the limits of graphical error and subject to the subsequent addition of braking torques arising from space-harmonic magnetomotive forces, friction, windage, iron losses and stray losses of the machine.

The object of the paper is to demonstrate that the braking characteristics can be determined by a mathematical, rather than a graphical, method which takes full account of the secondary-circuit reactance. In consequence, the compilation of the characteristics is simplified, the possibility of error is decreased (particularly at high magnetizations), the conditions for maximum torque can be established quantitatively and a clearer concept of the effects of magnetic saturation can be obtained.

(3) THEORY

(3.1) Equivalent Circuit

It can be shown that the circuit in Fig. 1 is an exact equivalent circuit per phase of a polyphase induction motor operating under the steady-state conditions of the primary (or secondary) windings carrying a constant direct current, provided that the space-harmonic m.m.f.'s and the iron and stray losses of the machine are ignored. The equivalent circuit may be referred to either the primary or the secondary circuit of the machine, the latter alternative being used here.

The alternating current I_1 is electromagnetically equivalent to the direct current, I_d , in the primary winding; i.e. the fundamental component of the m.m.f. distribution in the air-gap has the same amplitude for the alternative conditions of a.c. and d.c. excitation of the primary winding. It is convenient to assume that the

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Butler is Reader in Electrical Machinery, University of Sheffield.

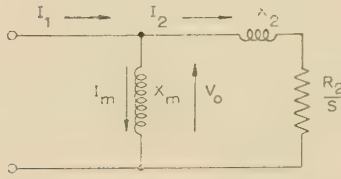


Fig. 1.—Equivalent circuit for d.c. braking conditions.

frequency of I_1 is the same as that during motoring, in which case the equivalent magnetizing reactance, X_m , of the primary winding and the total reactance, X_2 , of the secondary circuit are also referred to this frequency and have the same magnitude as that for motoring conditions, within the limits of the assumption that the effects of the space harmonics of m.m.f. be ignored.

The effective resistance of the secondary circuit at a rotor speed of N r.p.m. is R_2/S , where S is the ratio of N to the rated synchronous speed, N_s , of the motor, and R_2 is the ohmic resistance of the secondary circuit. Thus, S may be termed the fractional speed of the rotor, as distinct from the fractional slip represented by s in general usage in induction-motor theory.

The customary procedure in analysing the well-known equivalent circuit for motoring conditions has, in general, been followed closely in the analysis of the circuit shown in Fig. 1, even when magnetic saturation is present; i.e. expressions for the secondary current I_2 and the torque in 'synchronous watts', T , have been derived from the equivalent circuit in terms of either s or S as the appropriate independent variable. The disadvantage of such a procedure in the case of dynamic braking under conditions of high magnetic saturation is that the fractional speed itself is a complex function of the degree of magnetic saturation.

In what follows, the secondary current, fractional speed and torque are derived from Fig. 1 in terms of the magnetizing current I_m as the independent variable, in order that the open-circuit curve, or an appropriate simplification thereof, may be used in a direct mathematical analysis of the braking performance.

It should be noted that the equivalent circuit shown in Fig. 1 is valid for the d.c. dynamic braking of both asynchronous and synchronous polyphase machines, and also asynchronous single-phase motors. Thus, the present analysis is valid for any one of these machines, within the limits of the assumptions stated above.

(3.2) Equations for Secondary Current, Fractional Speed and Torque

It is shown in Section 9.1 that the secondary current in amperes, fractional speed, and torque in 'synchronous watts' can be expressed respectively as follows:

$$I_2 = \sqrt{\frac{I_1^2 - I_m^2}{1 + 2X_2/X_m}} \quad (1)$$

$$S = \frac{R_2}{\sqrt{\left(\frac{V_0^2}{I_2^2} - X_2^2\right)}} \quad (2)$$

and

$$T = mI_2^2 \frac{R_2}{S} \quad (3)$$

where m is the number of phases of the secondary (e.g. rotor) winding, V_0 is the induced e.m.f. per phase at the synchronous speed of the rotor and an equivalent magnetizing current I_m , and

$$X_m = \frac{V_0}{I_m} \quad (4)$$

That is, I_2 , S and T are defined in terms of I_m , V_0 and X_m for a fixed value of I_1 . Obviously, V_0 and X_m are defined directly in terms of I_m by the open-circuit characteristic of the machine.

Eqn. (1) shows that I_2 is independent of the manner in which the secondary resistance may vary with frequency and saturation conditions. Thus, provided that the variation of R_2 is known from experimental or design data, no difficulty is experienced in evaluating the braking characteristics of the machine from eqns. (1), (2) and (3). The deep-bar type of squirrel-cage rotor is of interest in this respect.

(3.3) Conditions for Maximum Torque

(3.3.1) Value of I_1 Known or Assumed.

It is shown in Section 9.2.1 that the maximum torque for a known value of I_1 occurs when V_0 , I_m and dV_0/dI_m have the values V_{0p} , I_{mp} and $(dV_0/dI_m)_p$ respectively at a point P on the open-circuit characteristic such that

$$\left(\frac{dV_0}{dI_m}\right)_p = \frac{(Z_{2p}^2 - 2X_2^2)(Z_{2p}^2 I_{mp} + X_2 V_{0p})}{(V_{0p} + 3X_2 I_{mp})Z_{2p}^2 - 2X_2^3 I_{mp}} \quad (5)$$

where

$$Z_{2p}^2 = \frac{V_{0p} + 2X_2 I_{mp}}{I_1^2 - I_{mp}^2} V_{0p} \quad (6)$$

Thus, when the point P has been established, the maximum torque and corresponding values of I_2 and R_2/S may be calculated from eqns. (3), (1) and (2) respectively.

(3.3.2) Value of I_m Known or Assumed.

It may be assumed that a particular point P on the open-circuit characteristic, at which V_0 , I_m and dV_0/dI_m have the known values of V_{0p} , I_{mp} and $(dV_0/dI_m)_p$, corresponds with the condition of maximum torque for a fixed value of I_1 yet to be evaluated. It is shown in Section 9.2.2 that in this case the value of I_2 is I_{2p} , for the condition of maximum torque, when

$$I_{2p}^2 = \frac{V_{0p}^2}{4X_2^2 V_a} [\sqrt{(V_b^2 + 8X_2 V_a I_{mp})} - V_b] \quad (7)$$

where

$$V_a = V_{0p} - I_{mp} (dV_0/dI_m)_p \quad (8)$$

and

$$V_b = \left(\frac{V_{0p}}{X_2} + 3I_{mp}\right) \left(\frac{dV_0}{dI_m}\right)_p - V_{0p} + 2X_2 I_{mp} \quad (9)$$

whilst the fixed value of I_1 is determined from

$$I_1^2 = \left(1 + \frac{2X_2 I_{mp}}{V_{0p}}\right) I_{2p}^2 + I_{mp}^2 \quad (10)$$

Thus, for any chosen value of $I_m = I_{mp}$, the maximum torque T_p and the corresponding values of $I_2 (=I_{2p})$, R_2/S and I_1 may be determined from eqns. (3), (7), (2) and (10), respectively.

It should be noted that $V_b^2 \gg 8X_2 V_a I_{mp}$ for small values of V_a , i.e. for only slight magnetic saturation, and consequently greater accuracy will be obtained for this case when eqn. (7) is rewritten more simply as

$$I_{2p}^2 = I_{mp}^2 \frac{V_{0p}^2}{X_2^2 V_b} \quad (11)$$

i.e. using the binomial expansion of the quantity under the root sign in eqn. (7). The error incurred in using eqn. (11) does not exceed 1% when $8X_2 V_a I_{mp}/V_b^2$ is not greater than 0.3.

Also, it may be noted that when magnetic saturation is absent under the maximum-torque conditions, V_a is zero and it is therefore inadmissible to divide by V_a in the derivation of eqn. (7). It is shown in Section 9.2.2 that in this case the well-known relationship $R_2/S = X_m + X_2$ is obtained as the condition for

maximum torque; the established theory for the case of no magnetic saturation must be utilized.

(3.3.3) Negligible Secondary Reactance.

When X_2 is so small that it may be neglected, eqn. (5) simplifies to

$$\left(\frac{dV_0}{dI_m}\right)_p = \frac{V_{0p}I_{mp}}{(I_1^2 - I_{mp}^2)} \quad (12)$$

Hence, when I_1 is known or assumed and $X_2 \approx 0$, eqn. (12) may be used to establish the maximum-torque conditions.

Alternatively, eqn. (12) may be rewritten as

$$I_{2p}^2 = \frac{V_{0p}I_{mp}}{\left(\frac{dV_0}{dI_m}\right)_p} \quad (13)$$

Thus, when I_{mp} is known or assumed and $X_2 \approx 0$, eqn. (13) may be used to establish the maximum-torque conditions.

(4) APPLICATION OF THEORY

In general, the open-circuit characteristic cannot be expressed accurately by a simple mathematical relationship between V_0 and I_m . Boyajian¹⁴ has shown that the replacement of the open-circuit or magnetization curve of an iron-cored reactor by two consecutive straight lines of different slope, to represent the unsaturated and saturated regions respectively, enables a mathematical solution to be obtained for various operating conditions of the reactor with useful analytical results. It is likely that results of general interest can be established in dynamic braking, by a similar procedure, particularly for transient conditions of operation. However, discussion here will be confined to the steady-state braking performance, and the actual open-circuit curve will be utilized.

(4.1) Open-Circuit Characteristic

The machine considered is that used by Harrison,¹³ i.e. a 15 h.p. 3-phase 400-volt 50 c/s 6-pole slip-ring induction motor. The secondary-winding (rotor) resistance and leakage reactance at 50 c/s are 0.27 and 0.56 ohm per phase, respectively. The transformation ratio of the effective windings per phase (primary/secondary) is 2.9.

Table 1 gives the open-circuit data of the machine obtained by measurement of the open-circuit slip-ring voltage and the direct current flowing to the star-connected stator winding via two of the three terminals. Since the direct current, I_d , for this

Table 1

OPEN-CIRCUIT DATA

I_1 and I_m , amp	5	10	15	20	30	40	50	60	70	80
V_0 , volts	54	110	145	163	185	198	207	214	220	225

connection, is related to the corresponding r.m.s. alternating current I_a by the relationship $I_a = 0.81I_d$, it follows that the alternating current referred to the secondary is $I_1 = 2.9I_a = 2.3I_d$. Hence, the value of I_1 corresponding to a direct stator current equal to the rated r.m.s. alternating stator current, of 42 amp, is 28 amp. Thus, the data in Table 1 are independent of the form of primary winding connection during braking, and give an upper limit corresponding to a direct stator excitation of approximately three times the rated current for the actual d.c. connection utilized.

(4.2) Maximum Torque and Critical-Speed Shift

The speed at which the maximum torque occurs, with a fixed value of R_2 , may be referred to as the critical speed, since unstable braking of the machine is obtained above this speed. Also, it is known⁴ that increased saturation due to an increased value of I_1 results in an increase of the critical speed, even though R_2 and X_2 are maintained constant. This effect may be conveniently referred to as the critical-speed shift.

It will be readily appreciated that the use of eqns. (7) and (11) is much less laborious than that of eqn. (5) for determining the maximum torques and critical-speed shifts for magnetic-saturation conditions corresponding to a succession of values of I_1 . Thus, the calculations may conveniently be divided into three stages corresponding to zero, slight and high magnetic saturation.

If the critical speed, as a fraction of the synchronous speed, is represented by S_c , the value of R_2/S_c for the region of no magnetic saturation in the characteristic represented by Table 1 may be taken as $11 + X_2$ up to a value of I_{mp} of 10 amp. Hence, the maximum torque in this region may be taken as

$$T_p = \frac{3I_{mp}^2}{\frac{1}{11 + X_2} + \frac{X_2}{60.5}} \quad (14)$$

whilst I_1 is obtained from

$$I_1^2 = \frac{2(11 + X_2)^2 I_{mp}^2}{121 + 2X_2(11 + X_2)} \quad (15)$$

as established in Section 9.2.3 for the region of no magnetic saturation.

Beyond this region, up to the point where I_{mp} is approximately equal to the rated current when X_2 is limited to the leakage reactance of the secondary winding, eqns. (11), (10), (2) and (3) may be used to determine I_{2p} , I_1 , R_2/S_c and T_p , respectively, for assumed successive values of I_{mp} . For higher values of magnetic saturation, eqn. (7) must be used in place of eqn. (11) for the determination of the value of I_{2p} for assumed successive values of I_{mp} .

The results for the machine in question are given in graphical form in Figs. 2 and 3. It will be seen that the alternative conditions of the secondary-circuit reactance being zero, one, five and ten times the secondary-winding leakage reactance have been considered. The latter two values of X_2 have been included in view of a recent tendency to incorporate components having

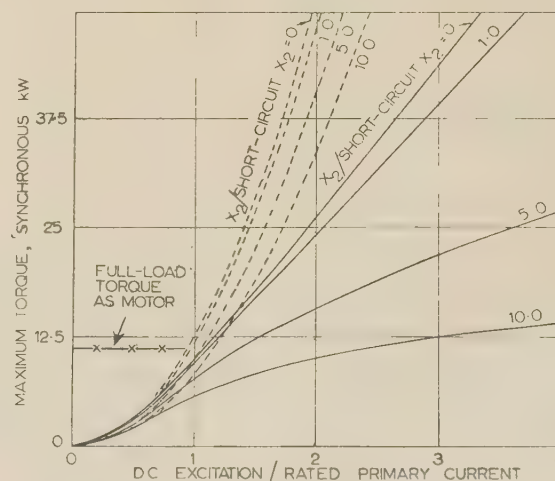


Fig. 2.—Variation of maximum torque with direct primary excitation.

----- Without magnetic saturation.
 ——— With magnetic saturation.

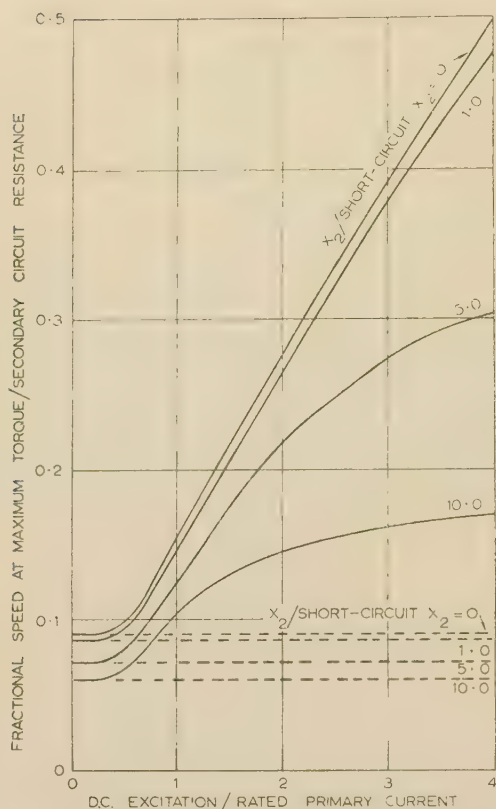


Fig. 3.—Variation of critical-speed/resistance ratio with direct primary excitation.

----- Without magnetic saturation.
 ——— With magnetic saturation.

appreciable inductive reactance in the external secondary circuit of mine winders, to facilitate automatic control of the braking torque. Although these two values of X_2 are excessive, the results are nevertheless of interest in stressing the deleterious effect of increased secondary-circuit reactance.

Evidently, for the particular case of zero secondary reactance, the results for the region of magnetic saturation are evaluated by using the simple relationship of eqns. (12) or (13), whilst eqns. (14) and (15) are used for the region of no magnetic saturation in their further amplified forms $T_p = 33I_{mp}^2$ and $I_1^2 = 2I_{mp}^2$.

(4.3) Torque/Speed Characteristic

The torque/speed characteristic is evaluated for fixed values of R_2 , X_2 and I_1 , by using successive values of I_m , and the corresponding values of V_0 and X_m , in conjunction with eqns. (1), (2) and (3). The number of successive values of I_m can be minimized by the availability of Figs. 2 and 3, and bearing in mind the values of I_{mp} used in compiling those curves. A typical example of torque/speed calculation is given as Table 2.

Obviously, for any given value of I_m , the torque expressed by eqn. (3) remains constant irrespective of the degree of magnetic saturation and variation of R_2 , provided the ratio R_2/S is maintained constant. Hence, when the torque/speed characteristic has been compiled for the lowest possible value of R_2 , the characteristic for any increased value will have the same maximum torque and can be readily compiled from the initial characteristic for the same range of speed by merely moving the characteristic forward, parallel to the speed axis, by the ratio of the second to the first value of R_2 . Thus, for each fixed value of I_1 it is sufficient to determine the single characteristic corresponding to $R_2 = 0.27$ ohm for the machine in question.

It may be noted here that the single characteristic corresponding to $R_2 = 0.27$ ohm is directly applicable to any greater value of R_2 merely by adding a further scale of S graduated in direct proportion to R_2 ; i.e., the characteristic for the lowest possible value of R_2 is inherently a characteristic of increased speed range for any increase of R_2 .

In order to obtain a clear conception of the effects of magnetic saturation on the torque/speed characteristics with different values of I_1 , it is helpful to use the 'ideal' characteristic as the basis of comparison, namely that obtained for the hypothetical condition of no magnetic saturation even at the largest values of I_1 . From Fig. 2 it is evident that the ideal maximum torque will be several times the actual torque at large values of I_1 .

It will be advantageous to express the torque in terms of per-unit ideal maximum torque, since the high-speed portions of the characteristics with different values of I_1 will then be coincident with the ideal characteristic, because I_m tends to be much less than I_1 at high speeds. Thus, the effects of magnetic saturation on the separate characteristics may be compared collectively.

Also, in order that the change due to saturation may be related to the continuous rating of the machine, the torque/speed curves have been evaluated from eqns. (1), (2) and (3) for values of I_1 equivalent to 1, 2 and 4 times the rated current flowing as direct current to the star-connected primary winding via two of the three terminals. It may be noted that this form of d.c. connection is usually preferred in practice, since it allows a simpler

Table 2

COMPILATION OF TORQUE/SPEED AND TORQUE/RESISTANCE CURVES FROM OPEN-CIRCUIT CURVE AND EQUIVALENT CIRCUIT
 PARAMETERS FOR $I_1 = 56$ AMP AND $X_2 = 0.56$ OHM

	Quantity	Load points									
		1	2	3	4	5	6	7	8	9	10
1	I_m	53	48	40	32	26	20	12			
2	V_0	210	206	198	188	176	164	127			
3	X_m	3.96	4.3	4.95	5.89	6.78	8.2	10.6	11		
4	I_2^2	255	664	1240	1770	2090	2400	2700		11	11
5	$1000S (=N)$	20.6	33.2	48.4	61	69.5	82.5	114	200	500	1000
6	T ('syn kW')	10.0	16.2	20.8	23.4	24.4	23.6	19.2	11.3	4.6	2.3
7	$1000R_2$	1310	816	558	443	389	327	234	135	54	27

Notes. (i) $R_2 = 0.27$ ohm for torque/speed curve in lines 6 and 5.

(ii) $S = 0.1$ for torque/resistance curve in lines 6 and 7.

(iii) Values of $1000R_2$ obtained directly from corresponding values of $1000S$ by using relationship $1000R_2 = 1000 \times 0.1 \times 0.27/S = 27000/1000S$.

(iv) Well-known relationship for constant X_m is used for last three load points, i.e. $T = 3X_m^2 I_1^2 R_2 / S [(R_2/S)^2 + (X_m + X_2)^2]$.

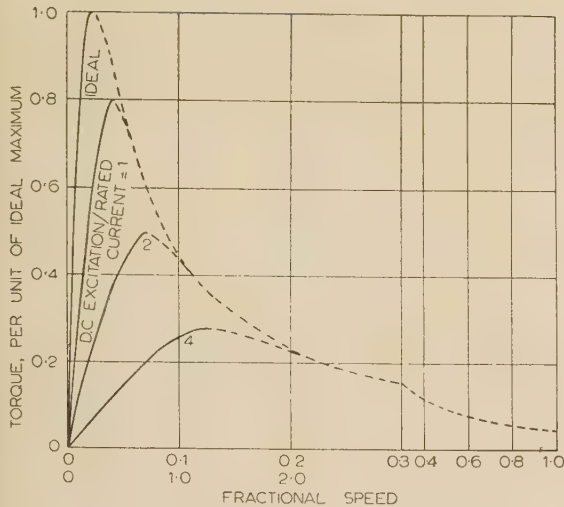


Fig. 4.—Torque/speed characteristics with X_2 at the short-circuit value.

— Stable.
 ---- Unstable.
 Upper scale of speed is for R_2 = short-circuit value.
 Lower scale of speed is for R_2 = $10 \times$ short-circuit value.

control scheme and a higher direct voltage to be used than is possible with the alternative connections available.

The results are shown in Fig. 4, from which it will be seen that even for the highest value of I_1 , magnetic saturation is absent above about 25% of the synchronous speed when no external resistance is included in the secondary circuit. It should be noted that the ideal torque/speed curve has been evaluated from

$$T = \frac{2}{Q + 1/Q} \quad \dots \quad (16)$$

where

$$Q = \frac{X_m + X_2}{R_2/S} \quad \dots \quad (17)$$

For the machine in question, $Q = 42.8S$. It should be further noted that a true comparison of the actual torques when I_1 is equivalent to 1, 2 and 4 times the rated current is obtained when the per-unit-of-ideal-peak-torque values are multiplied by 1, 4 and 16 respectively.

(4.4) Torque/Resistance Characteristic

In mine-winders and cranes, the problem of maintaining stable braking operation with variation of the secondary-circuit external resistance is more readily investigated by the availability of torque/resistance characteristics at fixed values of S and I_1 than by the availability of torque/speed characteristics at fixed values of R_2 and I_1 . The secondary-circuit resistance at which the maximum torque occurs may be referred to as the critical resistance, since the reduction of R_2 below its critical value, R_{2c} , results in a reduction of the braking torque with consequent instability of the braking operation.

Evidently, the torque/resistance characteristic may be evaluated by using successive values of I_m , and the corresponding values of V_b and X_m in conjunction with eqns. (1), (2) and (3). It will be appreciated from the latter equations that when the characteristic has been compiled for a low value of S , the characteristic for any increased value will have the same maximum torque and can be readily compiled from the initial characteristic for an increased range of R_2 by merely moving the characteristic forward, parallel to the resistance axis, by the ratio of the second to the first value of S . Thus, for each value of I_1 , it is sufficient

to determine the single characteristic corresponding, e.g. to $S = 0.1$.

As in the case of the torque/speed characteristics, the ideal characteristic may be used as the basis of comparison and the torques may be expressed in terms of per-unit ideal maximum torque, since the low-resistance portions of the characteristics with different values of I_1 will then be coincident with the ideal characteristic, because I_m tends to be much less than I_1 at low resistance values. Thus the effects of magnetic saturation on the separate characteristics may be compared collectively.

The results are shown in Fig. 5. The ideal characteristic has been evaluated from eqn. (16), with $Q = 1.156/R_2$ for the machine in question. It will be seen that alternative resistance scales have been included in Fig. 5, since, instead of moving the

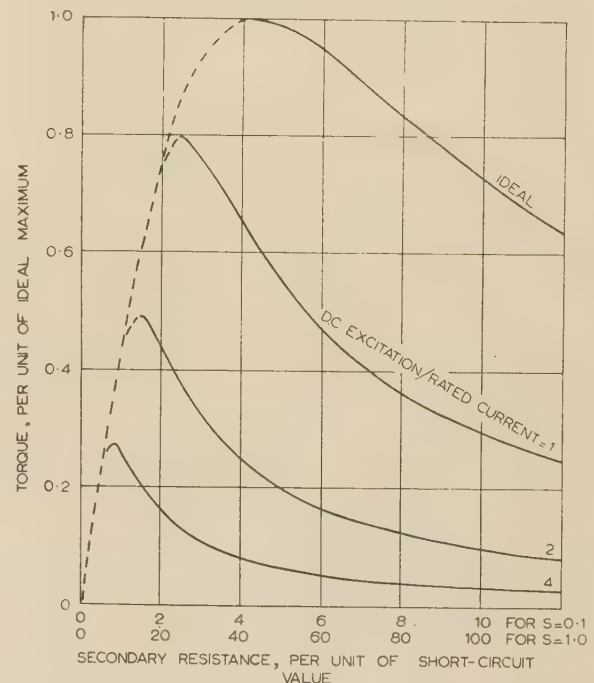


Fig. 5.—Torque/resistance characteristics with X_2 at the short-circuit value.

— Stable.
 ---- Unstable.

characteristics forward for increased values of speed, as discussed above, it may be more convenient to use the same characteristic to illustrate the performance at increased speeds. Also, it may be noted here that the torque/resistance characteristics may be compiled directly from the torque/speed characteristics, as discussed in detail in Section 9.3. A typical example of torque/resistance calculation is given in Table 2.

As in the case shown in Fig. 4, a true comparison of the actual torques when I_1 is equivalent to 1, 2 and 4 times the rated current is obtained when the per-unit-of-ideal-peak-torque values are multiplied by 1, 4 and 16, respectively. Fig. 6 has been compiled, since further useful information is more readily apparent by a comparison of the actual torques. However, it will be seen that the torques are here expressed in terms of per-unit maximum torque at rated current, as a more convenient basis of comparison of the actual values.

(5) CONFIRMATION OF THEORY

It has been confirmed that the present method gives agreement between the theoretical and experimental torque/speed and

torque/resistance characteristics which is at least as close as that obtained by Harrison,¹³ for the same machine.

(6) DISCUSSION OF RESULTS AND CONCLUSIONS

It will be seen from Fig. 2 that with low secondary reactance the maximum torque increases almost linearly with the direct primary current, except at low currents. Thus magnetic saturation results in a considerable decrease in the maximum torque below the ideal value, which varies as the square of the direct primary current. For example, when the direct stator current is 1, 2 and 4 times the rated current, the actual maximum torque is respectively 75, 50 and 27.5% of the ideal value. Increased secondary reactance appreciably worsens matters; with the foregoing values of direct stator current, the actual maximum torque with a secondary reactance of five times the short-circuit value is respectively 76, 60 and 34% of that with negligible secondary reactance.

Fig. 3 shows that with low secondary reactance the critical speed also increases almost linearly with the direct primary current, except for low currents. For the machine in question, the increase is approximately 125% for each 100% increase in the direct primary current above the rated value. The effect is somewhat reduced when the secondary reactance is increased.

Ignoring the secondary leakage reactance of the machine, as Stul'nikov¹⁰ and others^{11,12} have done, results in percentage positive errors of 6, 9 and 16 in the maximum torque (Fig. 2), and approximately 5, 5 and 5 in the critical speed (Fig. 3), for direct primary currents of 1, 2 and 4 times the rated current, respectively. Hence, the leakage reactance may be neglected for all practical values of the direct primary current and for most practical applications, so far as the critical speed (or critical resistance) is concerned, but not so far as the torque is concerned.

Figs. 4 and 5 show that the effect of magnetic saturation, even at rated current, is to distort the stable portions of the torque/speed and torque/resistance characteristics severely, whilst affecting the unstable portions much less. Thus, saturation results in the torque/speed and torque/resistance characteristics being appreciably different in shape; the ideal torque/speed and torque/resistance characteristics are identical in shape and are represented by a single characteristic when the resistance scale

is taken as $(X_m + X_2)$ times that of the speed scale (see Section 9.3).

It will be seen from Fig. 6 that although the maximum torque increases more than proportionately with the direct primary current, the increase of torque is much less than proportionate when R_2 is constant and greater than the critical resistance corresponding to the rated current. For example, as the direct primary current is successively 1, 2 and 4 times the rated value, the torques are only 1, 1.4 and 2 times when R_2 is twice the above critical resistance, and only 1, 1.2 and 1.4 times when R_2 is several times the critical resistance. Thus, in the case of a slip-ring induction motor, for which the braking torque is controlled by variation of the external secondary resistance, it is desirable that R_2 should approach closely to the appropriate critical value at all speeds if maximum benefit is to be obtained from increased direct primary currents. In a squirrel-cage motor applied to drives subject to frequent starting and stopping, it may be deduced from Fig. 4 that the braking performance is improved when the rotor resistance is increased beyond that of even the conventional high-resistance rotor.

(7) ACKNOWLEDGMENTS

The author gladly acknowledges the help received in discussions with Dr. D. Harrison. The facilities provided by the Electrical Engineering Department of the University of Sheffield are also acknowledged.

(8) REFERENCES

- (1) ROSENBERG, E., and PECK, J. S.: 'Improvements in Braking Induction Motors', British Patent Application No. 17010, 1910.
- (2) HELLMUND, R. E.: 'Bremsen von Induktionsmotoren mittels Gleichstrom', *Elektrotechnik und Maschinenbau*, 1910, **28**, p. 837.
- (3) HARRELL, F. E., and HOUGH, W. R.: 'D.C. Braking of Induction Motors', *Electrical Engineering*, 1935, **54**, p. 488.
- (4) WORRALL, R. W.: 'Dynamic Braking of A.C. Winders', *Metropolitan-Vickers Gazette*, 1938, **17**, p. 239.
- (5) LA PIERRE, W., and METAXAS, N.: 'D.C. Dynamic Braking of Squirrel-Cage Motors', *Transactions of the American I.E.E.*, 1953, **72**, p. 981.
- (6) ROSENBERRY, G. M.: Discussion of Reference 5, *ibid.*, p. 984.
- (7) COCHRAN, P. L.: 'Calculation of Dynamic Braking Characteristics of Wound Rotor Induction Motors', *ibid.*, p. 992.
- (8) MULLIGAN, J. W.: 'Dynamic Braking of Slipring Induction Motors Applied to Mine Winders', *Metropolitan-Vickers Gazette*, 1953, **24**, p. 375.
- (9) EVERT, C. F.: 'Dynamic Braking of Squirrel-Cage Induction Motors', *Transactions of the American I.E.E.*, 1954, **73**, p. 242.
- (10) STUL'NIKOV, V. I.: 'Calculation of the Maximum Torque and Critical Slip in the Dynamic Braking of an Induction Motor', *Elektrichestvo*, 1955, No. 9, p. 15.
- (11) WEISSHEIMER, H.: 'Die Asynchronmaschine mit Gleichstrom erregt als Bremse', *Archiv für Elektrotechnik*, 1934, **28**, p. 487.
- (12) FRIEDLANDER, E.: 'Principle and Features of a New Dynamic Braking and Motor Control System for A.C. Winders', *G.E.C. Journal*, 1949, **16**, p. 204.
- (13) HARRISON, D.: 'The Dynamic Braking of Induction Motors', *Proceedings I.E.E.*, Paper No. 1885 U, August, 1955 (103 A, p. 121).
- (14) BOYAJIAN, A.: 'Mathematical Analysis of Non-Linear Circuits', *General Electric Review*, 1931, **34**, pp. 531 and 745.

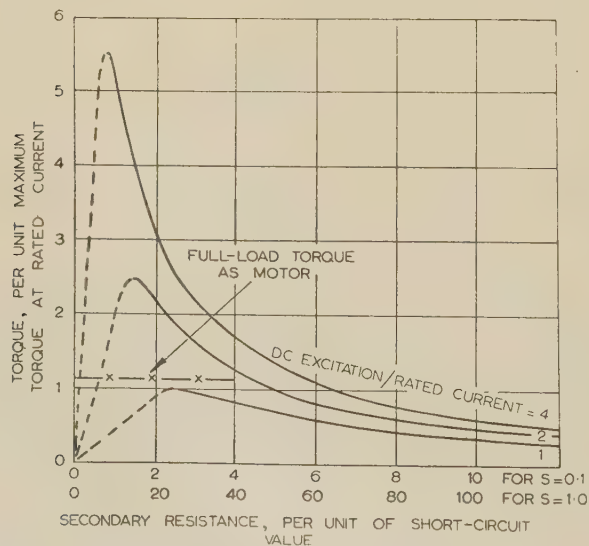


Fig. 6.—Torque/resistance characteristics with X_2 at the short-circuit value.

— Stable.
--- Unstable.

(9) APPENDICES

(9.1) Secondary Current, Speed and Torque as Functions of I_m

From the equivalent circuit shown in Fig. 1,

$$V_0^2 = I_2^2 \left(X_2^2 + \frac{R_2^2}{S^2} \right)$$

from which

$$S = \frac{R_2}{\sqrt{\left(\frac{V_0^2}{I_2^2} - X_2^2 \right)}} \quad (2)$$

Also,

$$I_m^2 = V_0^2 / X_m^2$$

and

$$I_1^2 = \frac{(R_2/S)^2 + (X_m + X_2)^2}{X_m^2} I_2^2$$

i.e. using eqn. (2),

$$I_1^2 = \frac{V_0^2 - (X_2 I_2)^2 + (X_m + X_2)^2 I_2^2}{X_m^2}$$

$$= \left(\frac{V_0}{X_m} \right)^2 + \left(1 + \frac{2X_2}{X_m} \right) I_2^2$$

$$= I_m^2 + \left(1 + \frac{2X_2}{X_m} \right) I_2^2$$

Thus

$$I_2 = \sqrt{\left(\frac{I_1^2 - I_m^2}{1 + \frac{2X_2}{X_m}} \right)}$$

whilst the torque in 'synchronous watts' is

$$T = \frac{m I_2^2 R_2}{S}$$

(9.2) Conditions for Maximum Torque

(9.2.1) Value of I_1 Known or Assumed.

Since the torque, T , is always of the same algebraical sign, the condition for maximum value of T is identical with that for maximum value of T^2 . Thus, from eqns. (3) and (2),

$$\left(\frac{T}{m} \right)^2 = I_2^4 \left(\frac{V_0^2}{I_2^2} - X_2^2 \right) = V_0^2 I_2^2 - X_2^2 I_2^4$$

Hence

$$\frac{1}{m^2} \frac{dT^2}{dI_m} = 2V_0 I_2^2 \frac{dV_0}{dI_m} + V_0^2 \frac{dI_2^2}{dI_m} - 2X_2^2 I_2^2 \frac{dI_2^2}{dI_m}$$

and, when T has its maximum value,

$$\frac{dV_0}{dI_m} = \left(\frac{X_2^2}{V_0} - \frac{V_0}{2I_2^2} \right) \frac{dI_2^2}{dI_m}$$

but, from eqn. (1),

$$I_2^2 = \frac{I_1^2 - I_m^2}{1 + \frac{2X_2}{X_m}}$$

Hence

$$\frac{dI_2^2}{dI_m} = - \frac{2I_m + \frac{2X_2 I_2^2 (V_0 - I_m \frac{dV_0}{dI_m})}{V_0^2}}{1 + \frac{2X_2}{X_m}}$$

Thus,

$$\frac{dV_0}{dI_m} = \left(\frac{V_0}{I_2^2} - \frac{2X_2^2}{V_0} \right) \frac{I_m + \frac{X_2 I_2^2 (V_0 - I_m \frac{dV_0}{dI_m})}{V_0^2}}{1 + \frac{2X_2 I_m}{V_0}}$$

i.e.

$$V_0^3 I_2^2 \left(1 + \frac{2X_2 I_m}{V_0} \right) \frac{dV_0}{dI_m} = (V_0^2 - 2X_2^2 I_2^2) \left[V_0^2 I_m + X_2 I_2^2 \left(V_0 - I_m \frac{dV_0}{dI_m} \right) \right]$$

$$\frac{dV_0}{dI_m} = \frac{(V_0^2 - 2X_2^2 I_2^2)(V_0 I_m + X_2 I_2^2 V_0)}{V_0^2 I_2^2 (V_0 + 2X_2 I_m) + (V_0^2 - 2X_2^2 I_2^2) X_2 I_2^2 I_m} = \frac{(V_0^2 - 2X_2^2 I_2^2)(V_0 I_m + X_2 I_2^2 V_0)}{(V_0 + 3X_2 I_m) V_0^2 I_2^2 - 2X_2^3 I_m I_2^4} \quad (18)$$

or

$$\frac{dV_0}{dI_m} = \frac{(Z_2^2 - 2X_2^2)(Z_2^2 I_m + X_2 V_0)}{(V_0 + 3X_2 I_m) Z_2^2 - 2X_2^3 I_m} \quad (19)$$

where

$$Z_2^2 = \frac{V_0^2}{I_2^2} = \frac{V_0(V_0 + 2X_2 I_m)}{I_1^2 - I_m^2} \quad (20)$$

(9.2.2) Value of I_m Known or Assumed.

From eqn. (18)

$$2X_2^3 V_a I_a^4 + X_2 V_0^2 V_b I_2^2 - V_0^4 I_m = 0 \quad (21)$$

where

$$V_a = V_0 - I_m \frac{dV_0}{dI_m} \quad (22)$$

and

$$V_b = \left(\frac{V_0}{X_2} + 3I_m \right) \frac{dV_0}{dI_m} - V_0 + 2X_2 I_m \quad (23)$$

Since I_2^2 must always be a positive quantity, the only admissible solution of eqn. (21) is

$$I_2^2 = - \frac{V_0^2 V_b}{4X_2^2 V_a} + \frac{\sqrt{(X_2^2 V_0^4 V_b^2 + 8X_2^3 V_a V_0^4 I_m)}}{4X_2^3 V_a} - \frac{(V_0/X_2)^2}{4V_a} [\sqrt{(V_b^2 + 8X_2 V_a I_m)} - V_b] \quad (24)$$

The corresponding value of I_1 in terms of I_2 and I_m is, from eqns. (1) and (4),

$$I_1^2 = \left(1 + \frac{2X_2 I_m}{V_0} \right) I_2^2 + I_m^2 \quad (25)$$

It should be noted that

$$V_0 = I_m \frac{dV_0}{dI_m}$$

when saturation is absent, i.e. $V_a = 0$, and it is therefore inadmissible to divide by V_a in deriving eqn. (24). Thus, when saturation is absent, it is necessary to substitute $V_a = 0$ in eqn. (21), in which case the well-known relationship of $R_2/S = X_m + X_2$ is obtained as the condition of maximum torque, bearing in mind that $V_0^2/I_2^2 = (R_2/S)^2 + X_2^2$.

(9.2.3) Region of No Magnetic Saturation.

From the equivalent circuit shown in Fig. 1,

$$X_m^2 I_m^2 = \left(X_2^2 + \frac{R_2^2}{S^2} \right) I_2^2 = \frac{\left(X_2^2 + \frac{R_2^2}{S^2} \right) X_m^2 I_1^2}{(X_m + X_2)^2 + \frac{R_2^2}{S^2}} \quad (26)$$

It is well known that for the case of no magnetic saturation, the maximum torque occurs when $R_2/S = X_m + X_2$. Hence, substituting in eqn. (26), the value of I_1 when the torque is a maximum is given by

$$I_1^2 = \frac{2(X_m + X_2)^2 I_m^2}{X_m^2 + 2X_2(X_m + X_2)} \quad (27)$$

Again, it is well known that for the case of no magnetic saturation the maximum torque is

$$T_p = \frac{mX_m^2 I_1^2}{2(X_m + X_2)}$$

and substituting for I_1^2 ,

$$T_p = \frac{mI_m^2}{\frac{1}{X_m + X_2} + \frac{2X_2}{X_m^2}} \quad (28)$$

Eqns. (14) and (15) are obtained from eqns. (28) and (27) for $m = 3$ and $X_m = 11$.

(9.3) Derivation of T/R_2 Curve from T/S Curve

It will be seen from Fig. 1 that for a given value of I_1 , irrespective of the degree of saturation, the torque is a function of the variable R_2/S , i.e.

$$T = f\left(\frac{R_2}{S}\right) \quad (29)$$

When R_2 is constant and equal, e.g. to R_{2a} , T is a function of S and is given by

$$T_a = f\left(\frac{R_{2a}}{S}\right) \quad (30)$$

Alternatively, when S is constant and, e.g. equal to S_1 , T is a function of R_2 and is given by

$$T_1 = f\left(\frac{R_2}{S_1}\right) \quad (31)$$

Hence, T_a and T_1 have identical values when $R_{2a}/S = R_2/S_1$, i.e.

$$R_2 = \frac{S_1 R_{2a}}{S} \quad (32)$$

or

$$S = \frac{S_1 R_{2a}}{R_2} \quad (33)$$

Thus, for a given value of I_1 , a particular value of torque on the T/R_2 characteristic, for $S = S_1$, is identical with that on the T/S characteristic, for $R_2 = R_{2a}$, when the corresponding value of S of the T/S curve is converted to the corresponding value of R_2 of the T/R_2 curve by using eqn. (32).

It will be seen that in performing this conversion, the rising (or falling) portion of the T/S characteristic at the lower (or higher) values of S becomes the falling (or rising) portion of the T/R_2 characteristic at the higher (or lower) values of R_2 . Hence, since the effect of saturation is mainly evident on the rising portion of the T/S characteristic, it follows that the effect of saturation will be mainly evident on the falling portion of the T/R_2 characteristic.

In the particular case of no magnetic saturation, the T/S characteristic represented by eqn. (30) may be written more specifically as

$$T_a = \frac{2T_p}{\frac{S}{S_c} + \frac{S_c}{S}} \quad (34)$$

whilst the T/R_2 characteristic represented by eqn. (31) may be written more specifically as

$$T_1 = \frac{2T_p}{\frac{R_2}{R_{2c}} + \frac{R_{2c}}{R_2}} \quad (35)$$

where

$$S_c = \frac{R_{2a}}{X_m + X_2}$$

and

$$R_{2c} = (X_m + X_2)S_1$$

Thus, the characteristics represented by eqns. (34) and (35) are identical when $S/S_c = R_2/R_{2c}$, i.e.

$$R_2 = (X_m + X_2)S \quad (36)$$

i.e. a T/S characteristic becomes a T/R_1 characteristic when the speed scale of the former is multiplied by $X_m + X_2$ to give the resistance scale of the latter.

THE DESCRIBING-FUNCTION ANALYSIS OF A NON-LINEAR SERVO MECHANISM SUBJECTED TO STOCHASTIC SIGNALS AND NOISE

By P. N. NIKIFORUK, B.Sc., Ph.D., Graduate, and J. C. WEST, B.Sc., Ph.D., Associate Member.

(The paper was received 22nd March, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

A technique is presented for evaluating the response of a specific type of non-linear servo mechanism to random signals, to sinusoidal signals contaminated by noise and to random signals contaminated by noise. In particular, a second-order servo mechanism incorporating derivative-of-error stabilization and subjected to torque limitation is considered. Theoretical responses are evaluated for a wide range of parameters, and experimental verification is included.

The analysis presented is basically an extension of the describing-function technique and can be applied to any second-order servo mechanism incorporating, either by design or inherently, a simple amplitude-dependent frequency-insensitive non-linearity.

LIST OF PRINCIPAL SYMBOLS

- E_{in} = Normalized amplitude of sinusoidal signal input to the system.
 P_{0n} = Normalized amplitude of sinusoidal signal input to the non-linear element.
 I_{sn} = Normalized random signal input to the system.
 σ_{sn} = Normalized random signal input to the non-linear element.
 I_{Nn} = Normalized noise input to the system.
 σ_{Nn} = Normalized noise input to the non-linear element.
 $\pm V_c$ = Limits of linear operation.
 T_d, T, G_1 = Constants of system.

(1) INTRODUCTION

In the analysis of linear systems the principle of superposition allows the behaviour of a given system subjected to an arbitrary signal to be evaluated if the behaviour for a given signal is known. For this reason the performance specification of a linear servo mechanism is most commonly based on the response to sinusoidal or step-function inputs. Consideration of noise contaminating the signal has led to the statistical analysis of system behaviour subjected to random inputs¹⁻³ with the aim of optimizing system parameters to minimize the unwanted effects of the noise.

For systems containing non-linear elements the picture is entirely different; superposition no longer holds. Thus the response of a non-linear system, which will now be amplitude dependent, to a given input function does not define the system, and it is not possible to calculate the response to any other signal from the knowledge of the response to one signal. Much work⁴⁻⁹ has, however, been done on the step-function responses and steady-state frequency responses of simple non-linear systems, even though these do not formally enable the response for other types of signal to be predicted. It was primarily of interest to see how the non-linearity in a system modified the behaviour it would have if linear. Step-function analysis has shown in particular how certain non-linear systems may have damped responses for signals below a certain level but become

unstable for step inputs above a critical value. Frequency-response analysis has led to the analytical explanation of 'jump' phenomena, sub-harmonic resonance, and the development of the describing-function technique.

It is still, however, of considerable importance (since a system in practice will in the main not be subjected to either step-function or steady sinusoidal inputs) to be able to specify non-linear system behaviour for the type of signal actually encountered. Thus consideration must be given to random, continuously varying input signals. Although experimental work^{8,10} on the response of simple non-linear systems to random signals has already been carried out, few mathematical techniques have been developed for such analysis.^{10,11}

In the paper an approximate technique for evaluating the response of a non-linear second-order servo mechanism containing one amplitude non-linear element is developed. Use is made of an extension¹⁰ to the describing function of a non-linear element for statistical analysis.

The inputs considered are all random but statistically stationary. Gaussian amplitude probability distribution has been taken, but the method can be applied for any probability distribution, provided it is known. Restriction to second-order systems is entirely for analytical simplicity; higher-order systems could be analysed, since this only increases the complexity of the 'linear' portion of the analysis. The signals considered are as follows:

- A stochastic signal defined by its power-density spectrum and amplitude-probability distribution.
- A sinusoidal signal with noise present at the input. This is of interest in relation to the effect of noise on the jump phenomenon for the sinusoidal output.
- A stochastic signal contaminated by noise which in this context is a second stochastic signal with a different spectrum but similar amplitude probability distribution.

(2) THE BASIC SYSTEM

The basic system to be considered is shown in Fig. 1. It is a simple remote-position-control servo mechanism using deriva-

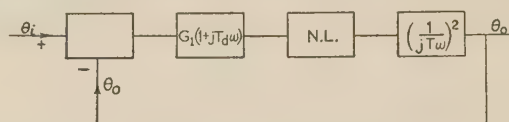


Fig. 1.—The basic system.

tive-of-error stabilization and incorporating a single-valued amplitude-dependent frequency-independent non-linearity. This system is of the 'anlis' class,* the harmonics, cross-modulation products and other distortion components generated by the non-linear element being severely attenuated by the filtering action^{4,8,10,11,13} of the servo motor.

(2.1) The Excitation Signals

The three types of input to be considered have already been mentioned in Section 1. The two most important characteristics

* This adjective is obtained from the initials of the phrase 'Amplitude Non-Linearity Integrator-Smoothed', which is sufficiently indicative of the class of systems implied.

Correspondence on Monographs is invited for consideration with a view to publication.
 Dr. Nikiforuk was formerly in the Electrical Engineering Department, University of Manchester, and is now with the Canadian Armaments Research and Development Establishment, Valcartier, Quebec.
 Dr. West is in the Electrical Engineering Department, University of Manchester.

of a random signal, and those which are required to describe it adequately, are the power spectrum and the amplitude probability distribution function. Random signals with Gaussian^{2,11,12,13} amplitude probability distributions occur most frequently in nature, and it is for this reason that this type of distribution is considered in the following analysis.

Random signal shaping by a non-linear element changes both the power spectrum and the amplitude probability distribution. In a closed-loop system the change in the latter, as far as the signal at the output of the system is concerned, is made negligible by the servo motor, the result of passing a non-Gaussian signal through a linear low-pass filter being that of making it more Gaussian.^{10,11,13}

(2.2) The Type of Non-Linearity Considered

The type of non-linearity that can be considered is any single-valued amplitude-dependent frequency-insensitive characteristic. Its response is characterized by the gain being solely dependent on the magnitude of the input. In particular, the saturation characteristic of Fig. 3 has been used, and in the analysis amplitudes of signals at various points in the system have been normalized with respect to V_c .

(3) THE CLOSED-LOOP RESPONSE

If a random signal with a Gaussian amplitude probability distribution is applied to a linear system, then every signal within the system is Gaussian. To a certain extent non-linearity within the closed-loop system destroys this nature of the signals. However, for the reasons stipulated in Section 2.1 it is assumed that the signal at the system output is Gaussian. For this condition the signal at the input to the non-linear element is also Gaussian, and this assumption facilitates the application of a quasi-linear representation of the non-linearity. In this¹⁰ the non-linear element is represented by an equivalent gain, the value of which depends upon the magnitude of the input to the non-linear element.

In the closed-loop system this equivalent gain and the input to the non-linear element must satisfy a condition that has yet to be derived. In addition, the r.m.s. value of this input must be that which results from an analysis of the system obtained when the non-linearity is replaced by the appropriate equivalent gain. The closed-loop response is evaluated from a simultaneous solution of these conditions.¹⁵

(3.1) The Equivalent Gain

There are two ways of representing the response of a non-linearity subjected to any type of input. The first is the 'single-function' representation. In this, a non-linear function of the form

$$x_0 = f(x_i) \quad (1)$$

is used to express the exact relationship between the instantaneous input x_i and the instantaneous response x_0 . The second is the 'two-function' representation. In this the non-linear element is represented by a quasi-linear amplifier and a distortion or harmonic generator. Thus, if the input is sinusoidal, this interpretation considers the response as the sum of a fundamental component and a distortion component consisting only of harmonics. For this second interpretation the response is represented by the expression

$$x_0 = k_{eg}x_i + x_H \quad (2)$$

where k_{eg} is the equivalent gain and is a function of signal amplitude and x_H is the distortion component. The quantity k_{eg} is known as the describing function for the non-linearity $f(x_i)$.

The basic assumption of the analysis is that in the closed-loop system the component x_H may be neglected for an appropriate choice of k_{eg} , which minimizes x_H . The k_{eg} function remains to be determined.

It has already been shown¹⁰ that the describing function may be interpreted as that linear gain k_{eg} which minimizes the mean-square harmonic content x_H in eqn. (2). That this assumption is the same as that made for the describing function follows from the fact¹² that the amplitude of each term in the Fourier series expansion of a function is such as to minimize the mean-square difference between that component and the function. Thus for a sinusoidal input k_{eg} is the gain relating the amplitude of the fundamental component of the output waveform to that of the input sinusoid.

From eqns. (1) and (2) the square of the instantaneous value of the distortion is

$$x_H^2 = [f(x_i) - k_{eg}x_i]^2$$

If $P_1(x_i)\delta x_i$ is the probability that x_i lies between x_i and $x_i + \delta x_i$ and is referred to as the first probability density function of x_i , then the mean-square value M of x_H^2 is

$$M = \int_{-\infty}^{\infty} [f(x_i) - k_{eg}x_i]^2 P_1(x_i) dx$$

The value of k_{eg} which minimizes M is given by

$$\begin{aligned} \frac{\partial M}{\partial k_{eg}} &= 0 \\ &= - \int_{-\infty}^{\infty} 2x_i [f(x_i) - k_{eg}x_i] P_1(x_i) dx \end{aligned}$$

On expansion

$$\int_{-\infty}^{\infty} f(x_i)x_i P_1(x_i) dx_i = k_{eg} \int_{-\infty}^{\infty} x_i^2 P_1(x_i) dx_i$$

Thus

$$k_{eg} = \frac{\int_{-\infty}^{\infty} f(x_i)x_i P_1(x_i) dx_i}{\int_{-\infty}^{\infty} x_i^2 P_1(x_i) dx_i} \quad (3)$$

This may be evaluated for particular functions of probability distribution $P_1(x_i)$ and non-linear characteristic $f(x_i)$.

(3.2) The Signal at the Input to the Non-Linear Element

It is well known in linear network analysis^{3,14} that the voltage spectrum of the output $\Phi_o(j\omega)$ is related to the voltage spectrum of the input $\Phi_i(j\omega)$ by the steady-state frequency response function $H(j\omega)$ of the system

$$\Phi_o(j\omega) = H(j\omega)\Phi_i(j\omega) \quad (4)$$

The output-power density spectrum $G_{op}(\omega)$ is also related to the input-power density spectrum $G_{ip}(\omega)$ by

$$G_{op}(\omega) = |H(j\omega)|^2 G_{ip}(\omega) \quad (5)$$

The mean value of the input power is

$$\overline{|\theta_i|^2} = \int_{-\infty}^{\infty} G_{ip}(\omega) d\omega$$

and that for the output

$$\overline{|\theta_o|^2} = \int_{-\infty}^{\infty} |H(j\omega)|^2 G_{ip}(\omega) d\omega \quad (6)$$

Let the input signal to the whole system have a fairly typical spectrum, i.e. flat over a certain bandwidth, and let it be given by the quadratic form

$$\Phi_{is}(j\omega) = \Phi_d \frac{1}{p^2 T_f^2 + 2pT_f + 1} \quad (7)$$

where $p = j\omega$ and Φ_d is the zero-frequency spectral density. Then the mean-square value of the input amplitude or mean power is

$$\bar{I}_s^2 = \int_{-\infty}^{\infty} \Phi_d^2 \left| \frac{1}{(1 - T_f^2 \omega^2) + j2T_f \omega} \right|^2 d\omega \quad (8)$$

$$\begin{aligned} &= \Phi_d^2 \int_{-\infty}^{\infty} \frac{d\omega}{(1 + T_f^2 \omega^2)^2} \\ &= \frac{\pi}{2T_f} \Phi_d^2 \quad (9) \end{aligned}$$

Φ_d^2 is the zero-frequency power spectral density.

General forms of the integral of eqn. (8) are evaluated in Reference 3, and those used in this paper are tabulated in Section 11.1.

If the non-linear element of Fig. 1 is replaced by the approximate linear gain of the describing function k_{eg} the closed-loop transfer function is given by

$$\theta_o = k_{eg} G_1 \frac{(1 + pT_d)}{T^2 p^2} (\theta_i - \theta_o)$$

and the signal at the input to the non-linearity is

$$G_1(1 + pT_d)(\theta_i - \theta_o) = \frac{G_1(1 + pT_d)p^2 T^2}{p^2 T^2 + G_1 k_{eg}(1 + pT_d)} \theta_i \quad (10)$$

Hence the mean-square input to the non-linearity is, from eqns. (7) and (8),

$$\begin{aligned} \sigma_s^2 &= \Phi_d^2 \int_{-\infty}^{\infty} \left| \frac{G_1(1 + pT_d)T^2 p^2}{[p^2 T^2 + G_1 k_{eg}(1 + pT_d)][p^2 T_f^2 + 2pT_f + 1]} \right|^2 d\omega \\ &= \Phi_d^2 \int_{-\infty}^{\infty} \left| \frac{C_2 p^2 + C_3 p^3}{d_0 + d_1 p + d_2 p^2 + d_3 p^3 + d_4 p^4} \right|^2 d\omega \\ &= 2\pi \Phi_d^2 \left[\frac{C_3^2(d_0 d_1 d_2 - d_0^2 d_3) + C_2^2 d_0 d_1 d_4}{2d_0 d_4(-d_0 d_3^2 + d_1^2 d_4 + d_1 d_2 d_3)} \right] \quad (11) \end{aligned}$$

where $C_2 = G_1 T^2$

$$C_3 = G_1 T_d T^2$$

$$d_0 = G_1 k_{eg}$$

$$d_1 = G_1 k_{eg}(T_d + 2T_f)$$

$$d_2 = T^2 + G_1 k_{eg}(T_f^2 + 2T_f T_d)$$

$$d_3 = 2T_f T^2 + G_1 k_{eg} T_d T_f^2$$

$$d_4 = T^2 T_f^2$$

For a given system with fixed values of T , T_d and G the expression in brackets in eqn. (11) is a function of the equivalent linear gain k_{eg} and the factor T_f determining the bandwidth of the input signal.

Hence, from eqn. (9), eqn. (11) can be written

$$\sigma_s^2 = 4T_f \bar{I}_s^2 U(T_f, k_{eg})$$

For convenience in dealing with a specific non-linearity such as that shown in Fig. 3, it is preferable to normalize the mean power input to the non-linearity σ_s^2 and the mean power of the input signal \bar{I}_s^2 with respect to the signal level at which linearity of operation ceases, i.e. V_c .

Thus

$$\sigma_{sn} = \sigma_s / V_c$$

and

$$\sqrt{(\bar{I}_{sn}^2)} = \sqrt{(\bar{I}_s^2)} / V_c$$

The equation relating these two powers then becomes

$$\sigma_{sn}^2 = 4T_f \bar{I}_{sn}^2 U(T_f, k_{eg}) \quad (12)$$

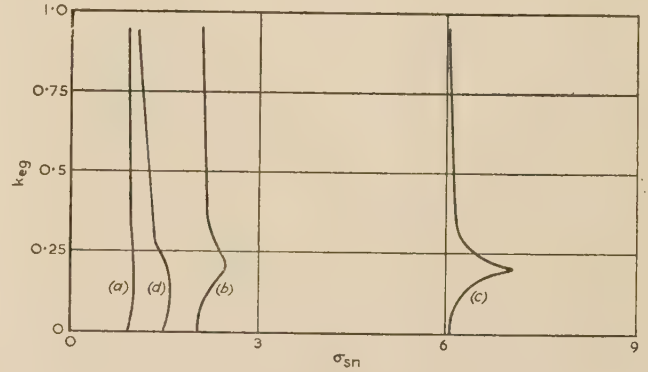


Fig. 2.—The input loci for random signals.

$$T_d = 4.27 \times 10^{-5} \text{ sec}$$

$$T = 7.45 \times 10^{-5} \text{ sec}$$

$$G_1 = 0.440$$

$$(a) \ I_{sn} = 0.71, T_f = 1.59 \times 10^{-5} \text{ sec}$$

$$(b) \ I_{sn} = 1.60, T_f = 1.59 \times 10^{-5} \text{ sec}$$

$$(c) \ I_{sn} = 4.65, T_f = 1.59 \times 10^{-5} \text{ sec}$$

$$(d) \ I_{sn} = 1.60, T_f = 7.45 \times 10^{-5} \text{ sec}$$

This relationship is plotted in Fig. 2 with k_{eg} as ordinate and σ_{sn} as abscissa for fixed values of T_f and \bar{I}_{sn}^2 . These particular axes are chosen to be of use in conjunction with the describing function of a particular non-linearity dealt with in Section 5.

(3.3) The Input to the Non-Linear Element for a Sinusoidal Signal

If the system input signal is sinusoidal of peak magnitude $E_{in} V_c$ then in normalized form it can be written as

$$E_{in} \exp j\omega t$$

The input to the non-linear element will also be sinusoidal and of the same frequency if an equivalent gain k_{eg} is assumed for the non-linearity. Let the peak amplitude of this signal be P_{on} also in normalized terms.

Then, from eqn. (10),

$$P_{on} = E_{in} \left| \frac{G_1(1 + j\omega T_d)T^2 \omega^2}{T^2 \omega^2 - G_1 k_{eg}(1 + j\omega T_d)} \right| \quad (13)$$

and is dependent on frequency $\omega/2\pi$.

(4) A PARTICULAR NON-LINEAR ELEMENT

The particular non-linear characteristic investigated has been confined to that of torque limitation. The characteristic is of the form shown in Fig. 3, non-linear operation commencing sharply at the signal levels $\pm V_c$. This characteristic is defined by the equation

$$\left. \begin{aligned} f(x_i) &= -V_c \text{ for } x_i < -V_c \\ &= x_i \text{ for } -V_c \leq x_i \leq V_c \\ &= V_c \text{ for } x_i > V_c \end{aligned} \right\} \quad (14)$$

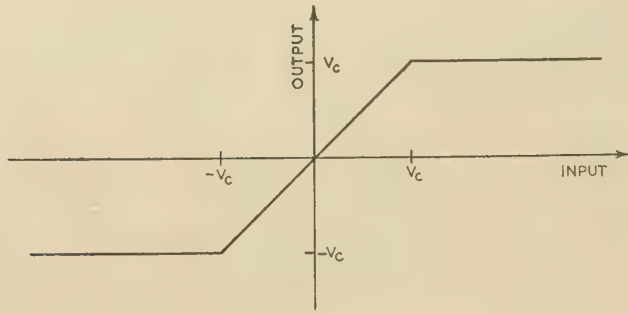


Fig. 3.—The non-linear characteristic.

(5) THE EQUIVALENT GAIN FOR RANDOM SIGNALS

If the excitation signal has a Gaussian distribution, then

$$P_1(x_i) = \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma_s^2}\right) \quad (15)$$

and the mean-square value of x_i is

$$\int_{-\infty}^{\infty} x_i^2 P(x_i) dx_i = \sigma_s^2$$

Hence to determine the equivalent gain for a non-linear function $f(x)$ eqn. (3) becomes

$$k_{eg} = \frac{1}{2\sigma_s^2} \int_{-\infty}^{\infty} x_i f(x_i) P_1(x_i) dx_i$$

From Appendix 11.2 this can be simplified to the form

$$k_{eg} = \frac{2}{\sqrt{2\pi}\sigma_s} \int_0^{V_c} \exp(-x_i^2/2\sigma_s^2) dx_i \quad (16)$$

for the case of saturation as shown in Fig. 3.

$$\text{Let } \sigma_{sn} = \sqrt{(2)\sigma_s/V_c} \quad (17)$$

$$\text{and } z = \frac{x_i}{\sigma_s \sqrt{2}} = x_i/V_c \sigma_{sn} \quad (18)$$

Then eqn. (16) becomes

$$k_{eg} = \frac{2}{\sqrt{\pi}} \int_0^{1/\sigma_{sn}} \exp(-z^2) dz \quad (19)$$

the solution of which is tabulated and written as

$$k_{eg} = \text{erf}\left(\frac{1}{\sigma_{sn}}\right) \quad (20)$$

The equivalent gain obtained by this expression is shown in Fig. 4 and may be called the 'gain locus'. In shape it is similar to the describing function for a sinusoidal input of varying amplitude, differing appreciably only in the region where non-linearity first sets in, i.e. in the vicinity of $\sigma_{sn} = 1$.

(5.1) The Closed-Loop Response to a Random Signal

The power spectrum of the input random signal is taken to be that given by eqn. (7), the r.m.s. value of which is given by eqn. (9). The r.m.s. value of the input to the non-linear element σ_{sn} is a function of k_{eg} for any non-linearity and is plotted from eqn. (12) in Fig. 2. The value of k_{eg} for a given σ_{sn} is shown in Fig. 4 for the particular non-linearity. In operation, both conditions must be satisfied. The appropriate values of equivalent

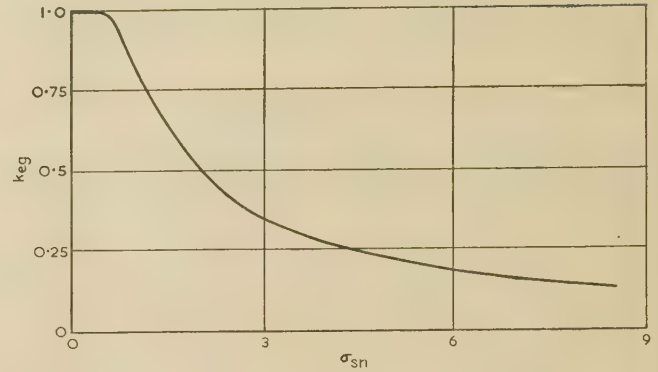


Fig. 4.—The equivalent gain for random signals.

gain for simultaneous solution are determined by the superposition of the input loci of Fig. 2 onto the gain locus of Fig. 4. The intersections of the appropriate loci determine the operating points of the system.

When the random signal is small the input locus is very near the k_{eg} -axis and the equivalent gain is approximately unity. The response of the non-linear system is approximately that obtained for linear-regime operation. As the signal is increased, the input locus moves in the direction of increasing σ_{sn} and the equivalent gain decreases. The non-linear response deviates, therefore, from linear-regime operation, the deviation increasing with increasing excitation signal.

Typical voltage spectra have been calculated in this manner and are shown in Fig. 5. Non-linear operation produces

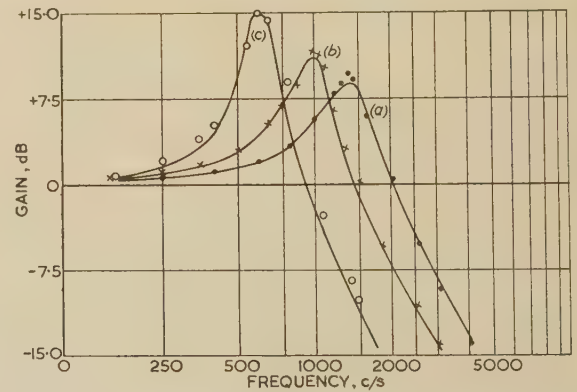


Fig. 5.—Typical responses to a random signal of constant bandwidth.

$$\begin{aligned} T_d &= 4.27 \times 10^{-5} \text{ sec} \\ T &= 7.45 \times 10^{-5} \text{ sec} \\ G_1 &= 0.440 \\ T_f &= 1.59 \times 10^{-5} \text{ sec} \end{aligned}$$

Theoretical responses are shown by line.
Experimental responses are shown by points.

- (a) Linear.
- (b) Non-linear; $I_{sn} = 1.60$.
- (c) Non-linear; $I_{sn} = 4.65$.

responses with decreased bandwidth and increased resonant peaks.

Eqn. (12) shows that the input loci are functions of signal bandwidth. This is shown graphically in Fig. 2, where the loci (b) and (d) are for excitation signals of the same r.m.s. value but of different bandwidth, the bandwidth of the signal corresponding to (b) being four times greater than that corresponding to (d). The manner in which this increase of gain due to signal-bandwidth limiting modifies the closed-loop response is shown in Fig. 6. The response (a) corresponds to the input locus (d) and the response (c) corresponds to the input locus (b).

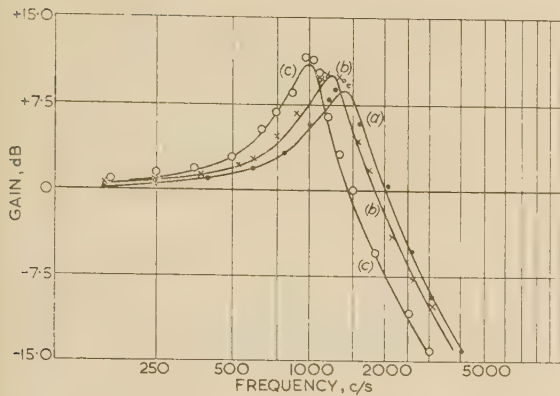


Fig. 6.—Typical responses to a random signal of varying bandwidth.

$$\begin{aligned} T_d &= 4.27 \times 10^{-5} \text{ sec} \\ T &= 7.45 \times 10^{-5} \text{ sec} \\ G_1 &= 0.440 \\ I_{sn} &= 1.60 \end{aligned}$$

Theoretical responses are shown by line.
Experimental responses are shown by points.

(a) Linear.
(b) Non-linear; $T_f = 7.45 \times 10^{-5} \text{ sec}$
(c) Non-linear; $T_f = 1.59 \times 10^{-5} \text{ sec}$

(5) THE EQUIVALENT GAIN FOR SINUSOIDAL SIGNALS CONTAMINATED BY NOISE

The equivalent gain, which is evaluated in Appendix 11.3, is a lengthy and tedious process. It is based on eqn. (3). First it is necessary to determine $P_1(x_i)$, the first probability density function for an input consisting of a sinusoidal signal, $P_0 \cos \omega t$ of mean power $W_s = \frac{1}{2}P_0^2$, together with a random noise with Gaussian probability distribution having a mean power $\sigma_N^2 = W_N$. This is given in eqn. (39).

Evaluation of the denominator of eqn. (3) gives the total mean power $\frac{1}{2}P_0^2 + W_N$. The equivalent gain, eqn. (50), has been evaluated for a wide range of parameters and a typical set of loci is shown in Fig. 7. If the noise component is zero the equivalent gain reduces to the describing function for sinusoidal input, as shown in Fig. 7(b), curve *a*, and if the signal is zero it reduces to that obtained for a random signal in Section 4.1. This is shown in Fig. 7(a), curve *a*. The input/output characteristic for saturation non-linearity is shown in Fig. 8.

(6.1) The Closed-Loop Response to a Sinusoidal Signal Contaminated by Noise

If the noise component of the input has a power spectrum of the form

$$\Phi_{iN} = \Phi_{oN} \left| \frac{1}{P_2 T_N^2 + 2PT_N + 1} \right|^2 \quad (21)$$

from an analysis similar to that given in Section 3.1 the r.m.s. value of the noise component of the input to the non-linear element is found to be

$$\sigma_N^2 = 2\pi \left[\frac{c_3^2(-d_0^2 d_3 + d_0 d_1 d_2) + c_2^2(d_0 d_1 d_4)}{2d_0 d_4(-d_0 d_3^2 - d_1^2 d_4 + d_1 d_2 d_3)} \right] \Phi_{oN} \quad (22)$$

where $c_2, c_3, d_0, \dots, d_4$ are defined by eqn. (11), except that T is replaced by T_N . The input loci represented by eqn. (22) are similar in form to those shown in Fig. 9.

The mean-square value of the signal component at the input to the non-linear element is, from eqn. (13),

$$P_0^2 = \left| \frac{G_1(1 + j\omega T_d)}{1 - G_1 k_{eg}(1 + j\omega T_d)/\omega^2 T^2} \right|^2 E^2 \quad (23)$$

The equivalent gain is a function of both signal and noise,

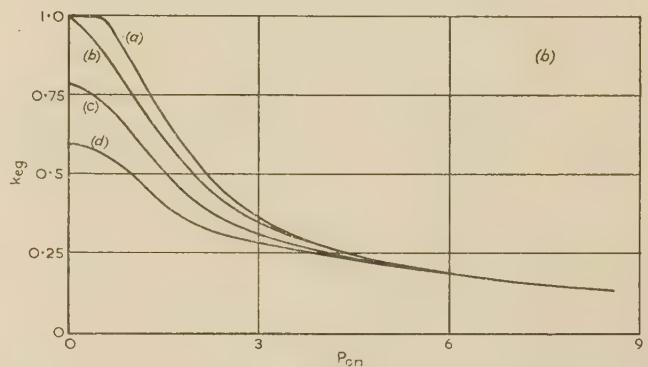
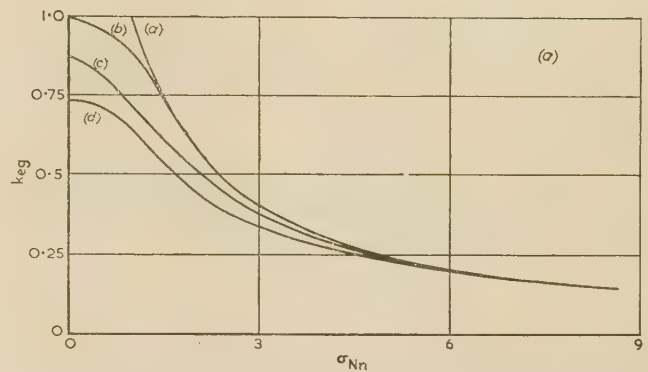


Fig. 7.—The equivalent gain for an input consisting of signal and noise.

(a) The gain loci for a constant noise component.

(a) $\sigma_{Nn} = 0.00$ (c) $\sigma_{Nn} = 0.94$
(b) $\sigma_{Nn} = 0.47$ (d) $\sigma_{Nn} = 1.33$

(b) The gain loci for a constant signal component.

(a) $P_{on} = 0.00$ (c) $P_{on} = 1.50$
(b) $P_{on} = 1.00$ (d) $P_{on} = 2.00$

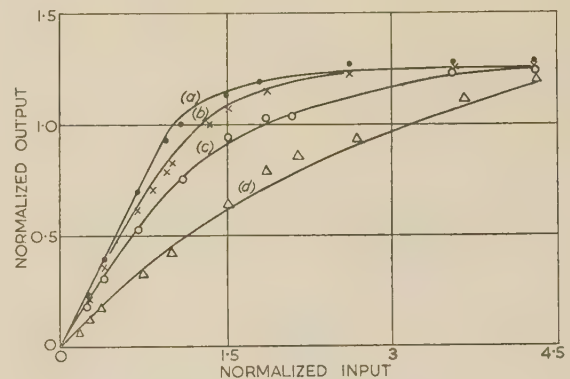


Fig. 8.—The response of the torque-limited element to a sinusoidal signal contaminated by noise.

Theoretical responses are shown by full line.
Experimental responses are shown by points.

(a) $\sigma_{Nn} = 0.00$ (c) $\sigma_{Nn} = 1.00$
(b) $\sigma_{Nn} = 0.50$ (d) $\sigma_{Nn} = 2.00$

and in the system the conditions imposed by eqn. (22) and (23) must be satisfied simultaneously with the equivalent gain of Fig. 7. The closed-loop response is evaluated from a simultaneous solution of these conditions. The technique employed is an extension of that developed in Reference 9 for purely sinusoidal inputs.

For the purpose of analysis let the loop remain closed, and let an input with a specified noise component and an arbitrary signal component be applied to the system. For such an input the relationship between k_{eg} and σ_N for an unknown P_{on} is

completely specified by eqn. (22) and is represented graphically by the appropriate input locus. Specific values of k_{eg} and σ_{Nn} are not yet known. However, since the signal component of the input is unspecified it is permissible to choose the amplitude and phase of this component of the error. In normalized form let the component be $E_{sn} \exp j\omega t$.

The stabilizing element is linear, and its response to the signal component is independent of the noise component. The signal component at the output of this element, and consequently at the input to the non-linear element, is therefore

$$AE_{sn} \exp j(\omega t + \phi)$$

where

$$A \exp j\phi = G_1(1 + j\omega T_d) \quad (24)$$

For this condition, σ_{Nn} unknown, the relationship between k_{eg} and P_{0n} is completely specified graphically by the appropriate gain locus of Fig. 7(b). The operating point of the system and the corresponding value of k_{eg} are obtained by superimposing the gain loci onto the input loci and determining the intersections of the appropriate loci. If the equivalent gain obtained in this manner is specified as k'_{eg} , the signal component at the output of the system is found to be

$$-\left(\frac{1}{\omega T}\right)^2 Ak'_{eg} E_{sn} \exp j(\omega t + \phi)$$

From the vector relationship

$$\theta_i = \theta + \theta_0$$

the input to the system is therefore

$$\theta_i = E_{sn} \exp j\omega t - \left(\frac{1}{\omega T}\right)^2 k'_{eg} AE_{sn} \exp j(\omega t + \phi) \quad (25)$$

In this equation angle θ_i is a function of ω , E_{sn} and I_{Nn} and is taken to be the representation of the open-loop frequency response. It is plotted in the complex plane as a function of the error, the reference vector. For an input consisting of a constant noise component and a signal component of constant frequency and variable amplitude, the resulting plot is a vector locus of the form shown in Fig. 9. Conversely, for a signal

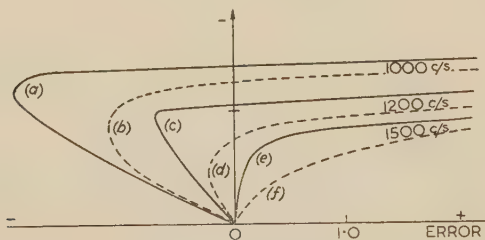


Fig. 9.—The open-loop vector loci.

(a), (c), (e) correspond to $\sigma_{Nn} = 0.00$
(b), (d), (f) correspond to $\sigma_{Nn} = 0.71$

component with a constant amplitude E_{in} the tip of the vector θ_i must lie on a circle of radius E_{in} , the centre being the origin. This latter plot is called the circle locus.

The procedure for evaluating the closed-loop response to the signal component is shown in Fig. 10. In this the signal has a constant amplitude and a variable frequency. At the very low frequency f_A there is but one intersection and hence only one operating point. As the frequency is increased the operating point moves uniformly along the circle locus in a clockwise direction. At the frequency f_B a condition is reached at which there are two intersections P_2 and P_4 . The latter is a tangency and is unstable for increasing signal frequency. Consequently

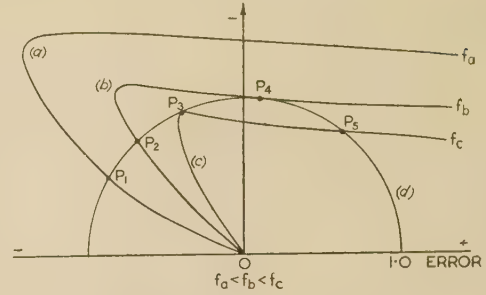


Fig. 10.—The evaluation of the closed-loop response to a sinusoidal signal contaminated by noise.

(a), (b) and (c) are vector loci.
(d) is a circle locus.

the operating point continues to move uniformly along the circle locus until the frequency f_C is attained. At this frequency the operating point P_3 is a tangency. At a slightly higher frequency there is but one intersection, and this is in the region of P_5 and at a much greater closed-loop gain. There is therefore a transition in the gain/frequency characteristic, the operating point moving discontinuously from P_3 to the region of P_5 . Following this transition the response for increasing signal frequency is uniform, the operating point continuing to move in a clockwise direction with increasing signal frequency.

If the signal frequency is not decreased the operating point moves uniformly along the circle locus in a counter-clockwise direction, passing through P_5 and continuing to P_4 , because P_3 , which is also a tangency, is unstable for decreasing frequencies. At P_4 the operating point is again a tangency and for a small decrease in frequency a transition takes place. The operating point moves discontinuously from P_4 to P_2 with a corresponding reduction in the closed-loop gain. The subsequent response for decreasing signal frequency is again uniform, the operating point continuing to move in a counter-clockwise direction.

Typical gain/frequency characteristics have been evaluated in this manner and are shown in Fig. 11. The effect of the noise

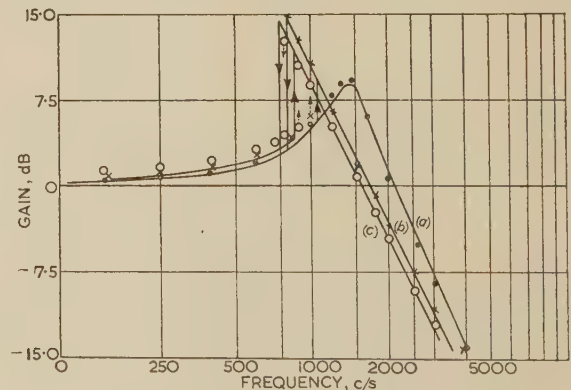


Fig. 11.—The response to a sinusoidal signal contaminated by noise of a constant bandwidth.

Theoretical responses are shown by line.
Experimental responses are shown by points.

$$T_d = 4.27 \times 10^{-5} \text{ sec}$$

$$T = 7.45 \times 10^{-5} \text{ sec}$$

$$G_1 = 0.440$$

$$T_N = 1.59 \times 10^{-5} \text{ sec}$$

(a) Linear

(b) Non-linear, $E_{in} = 1.89$, $I_{Nn} = 0.00$

(c) Non-linear, $E_{in} = 1.89$, $I_{Nn} = 0.71$

upon the response to the sinusoidal signal is twofold. First, it quenches the jump phenomena in the gain/frequency characteristic, the change in the upper transitional frequency being

much greater than that in the lower. Secondly, it increases the closed-loop gain at frequencies below resonance and decreases it at frequencies above resonance.

The effect on the input loci of limiting the noise bandwidth is similar to that described in Section 5.1 for random signals. The effect of this upon the closed-loop response to the signal is shown in Fig. 12, the effect of the noise becoming smaller with decreasing noise bandwidth.

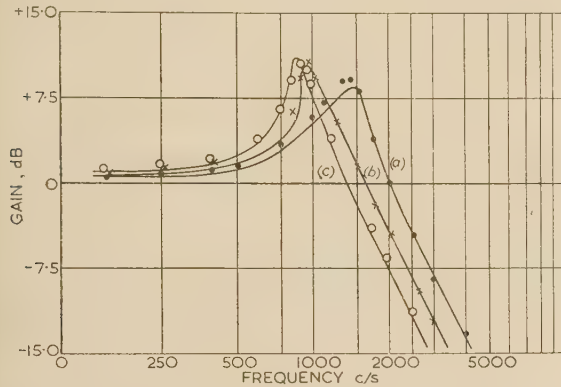


Fig. 12.—The response to a sinusoidal signal contaminated by noise of variable bandwidth.

Theoretical responses are shown by line.
Experimental responses are shown by points.

$$T_d = 4.27 \times 10^{-5} \text{ sec}$$

$$T = 7.45 \times 10^{-5} \text{ sec}$$

$$G_1 = 0.440$$

$$I_{Nn} = 1.41$$

(a) Linear

(b) Non-linear, $E_{in} = 1.89$, $T_N = 7.45 \times 10^{-5} \text{ sec}$

(c) Non-linear, $E_{in} = 1.89$, $T_N = 1.59 \times 10^{-5} \text{ sec}$

(7) THE EQUIVALENT GAIN FOR A RANDOM SIGNAL CONTAMINATED BY NOISE

If a random signal with a Gaussian amplitude probability distribution is contaminated by noise with a similar amplitude distribution, the first probability density function of the sum is, from Appendix 11.4,

$$Q(x_i) = \frac{1}{\sqrt{(2\pi)\sqrt{(\sigma_s^2 + \sigma_N^2)}}} \exp \left[-\frac{1}{2} x_i^2 / (\sigma_s^2 + \sigma_N^2) \right] \quad (26)$$

By substituting this expression together with eqn. (14) into eqn. (3), and by using an analysis similar to that given in Section 4.1, the latter equation is converted to

$$k_{eg} = \text{erf} \left[\frac{1}{\sqrt{(\sigma_{sn}^2 + \sigma_{Nn}^2)}} \right] \quad (27)$$

where σ_{sn} and σ_{Nn} have the same meaning as before.

This normalized gain locus is plotted in Fig. 13 only as a function of the noise component, the signal component being considered a parameter. The corresponding loci for the signal component with a constant noise component are precisely the same as those in Fig. 13, the appropriate changes in σ_{sn} and σ_{Nn} being taken into consideration.

(7.1) The Closed-Loop Response to a Random Signal Contaminated by Noise

Since the power spectrum of the signal is taken to be of the form given by eqn. (7) and that of the noise of the form given by eqn. (21), σ_s and σ_N are given by eqns. (12) and (22) respectively. The equivalent gain is a function of both signal and noise, and in the closed loop the conditions imposed by eqns. (12), (22) and (27) must be satisfied simultaneously. The closed-loop response is evaluated from a simultaneous solution of these

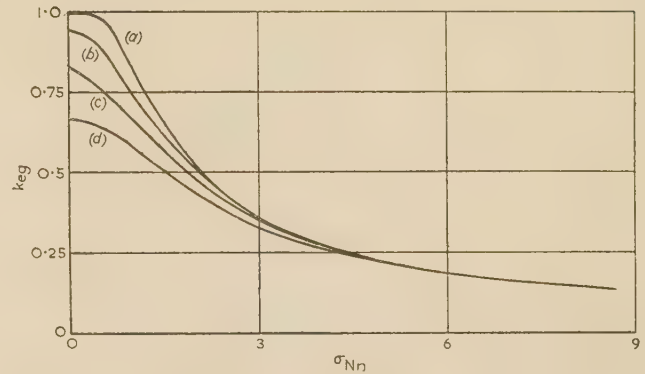


Fig. 13.—The equivalent gain-loci for an input consisting of random signal and noise.

$$(a) \sigma_{sn} = 0.00 \quad (c) \sigma_{sn} = 1.00$$

$$(b) \sigma_{sn} = 0.71 \quad (d) \sigma_{sn} = 1.42$$

conditions, and the technique employed is a combination of those described in Sections 5.1 and 6.1.

For the purpose of analysis, let the loop be closed and let an input with a specified noise component and arbitrary signal component be applied to the system. The first relationship to be considered is that between k_{eg} and σ_{Nn} . For the input specified, I_{sn} unknown, the relationship between k_{eg} and σ_{Nn} is defined by eqn. (22) and is represented graphically by a specific input locus. The specific values of k_{eg} and σ_{Nn} are still unknown but, nevertheless, must lie somewhere on this input locus. The second relationship to be considered is that between k_{eg} and σ_{sn} , which is represented graphically by the loci of Fig. 13.

To calculate the closed-loop response for a specified signal component the appropriate input locus is plotted on the equivalent gain-loci of Fig. 13. Since σ_{sn} is still unknown the operating point is yet undefined. To determine this point a particular value of k_{eg} is chosen. The point at which this intersects the gain loci is determined with the aid of the input locus and the corresponding value of σ_{sn} obtained. From this and eqn. (11) the corresponding value of I_{sn} is calculated. This computed value is compared with the specified value, and if the two should differ the calculation is repeated, the original choice of k_{eg} being modified in such a manner as to have the calculated value of I_{sn} correspond with the specified value. With practice this can be accomplished in three or four trials. This value of k_{eg} which equalizes the calculated and specified values of I_{sn} is the operating point of the system.

With the type of input considered the output-input gain-frequency characteristic has little meaning. The most implicit characteristic is the true error response, this being the difference between the desired output and the actual output. This characteristic has been evaluated in the manner described and is shown in Fig. 14.

(8) EXPERIMENTAL RESULTS

The experimental system was of the form shown by Fig. 1 and was simulated on an analogue computer. The non-linear characteristic was of the form shown in Fig. 3, non-linear operation commencing at ± 1.5 volts. Stabilization was achieved by derivative of error.

The response of the system to a random signal is shown in Figs. 5 and 6. There is good agreement with the predicted response.

The response of the non-linear element to a sinusoidal signal contaminated by noise is shown in Fig. 8. It is of the nature predicted and there is agreement with the calculated responses. The closed-loop response to the same input is shown in Fig. 11.

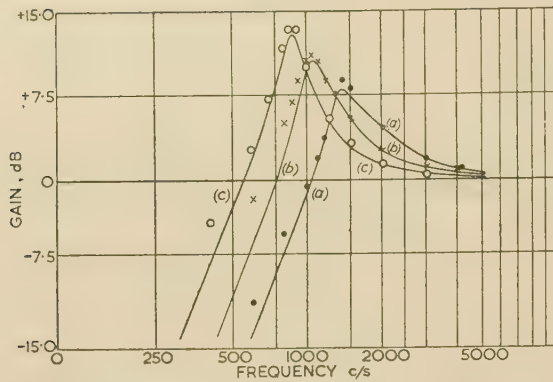


Fig. 14.—The true error response to a random signal contaminated by noise.

Theoretical responses are shown by line.
Experimental responses are shown by points.
 $T_d = 4.27 \times 10^{-5}$ sec $T_f = 3.18 \times 10^{-5}$ sec
 $T = 7.45 \times 10^{-5}$ sec $T_N = 0.792 \times 10^{-5}$ sec
 $G_1 = 0.44$
(a) Linear
(b) Non-linear, $\sigma_{sn} = 1.89$, $\sigma_{Nn} = 0.00$
(c) Non-linear, $\sigma_{sn} = 1.89$, $\sigma_{Nn} = 1.89$

Except for one small discrepancy the experimental and predicted responses are in agreement. The experimental responses show that the lower transitional frequency is unaffected by increasing noise power, whilst for the same condition the predicted responses show that this transitional frequency decreases. Fig. 12 shows that the effect of limiting the noise bandwidth is of the nature predicted.

The true error response is shown in Fig. 14. The agreement between the experimental and the predicted responses is not as good as would be expected from the previous comparisons. The greatest discrepancies occur at low frequency, and this can be attributed partially to the limitations of the equipment used at low error signals.

(9) CONCLUSIONS

A method has been presented for evaluating the response of non-linear servo mechanisms to random signals, to sinusoidal signals contaminated by noise, and to random signals contaminated by noise. Typical responses have been evaluated for a second-order system subjected to torque limitation. The system has been simulated on an analogue computer and experimental verification obtained.

It is concluded that the approximations made in this analysis are justified and, furthermore, that the techniques presented are sufficiently accurate to justify their application to other systems incorporating simple amplitude non-linearities.

(10) REFERENCES

- (1) PELEGRIN, M. J.: 'Application of Statistical Techniques to the Servomechanism Field, Automatic and Manual Control' (Butterworth, 1952), p. 123.
- (2) JONES, S.: 'The Determination of the Best Form of Response

for a Servo when an Extraneous Random Disturbance is Present with Error Signal', *ibid.*, p. 139.

- (3) JAMES, H. M., NICHOLS, N. B., and PHILLIPS, R. S.: 'Theory of Servo-Mechanisms' (McGraw-Hill, 1947), Chapter 8.
- (4) JOHNSON, E. C.: 'Sinusoidal Analysis of Feedback Control Systems containing Non-Linear Elements', Thesis (Sc.D.), Department of Electrical Engineering, M.I.T., 1951.
- (5) WEST, J. C., and NIKIFORUK, P. N.: 'The Behaviour of a Remote-Position-Control Servo Mechanism with Hard Spring Characteristics', *Proceedings I.E.E.*, Paper No. 1621 M, February, 1954 (101, Part II, p. 481).
- (6) WEST, J. C., and DOUCE, J. L.: 'The Frequency Response of a Certain Class of Non-Linear Feedback Systems', *British Journal of Applied Physics*, 1954, 5, p. 204.
- (7) LEVINSON, E.: 'Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System', *Applications and Industry* (American I.E.E.), March, 1953, No. 1, p. 1.
- (8) WEST, J. C., and NIKIFORUK, P. N.: 'The Response of Remote-Position-Control Systems with Hard-Spring Characteristics to Step-Function and Random Inputs', *Proceedings I.E.E.*, Paper No. 1728 M, October, 1954 (102 B, p. 575).
- (9) WEST, J. C., and NIKIFORUK, P. N.: 'The Frequency Response of a Servomechanism Designed for Optimum Transient Response', *Applications and Industry* (American I.E.E.), September, 1956, p. 234.
- (10) BOOTON, R. C.: 'Non-linear Control Systems with Statistical Inputs', Report No. 61, Dynamic Analysis and Control Laboratory, M.I.T.
- (11) BARRETT, J. F., and COALES, J. F.: 'An Introduction to the Analysis of Non-Linear Control Systems with Random Inputs', *Proceedings I.E.E.*, Monograph No. 154 M, November, 1955 (103 C, p. 190).
- (12) REDDICK, N. W., and MILLER, F. H.: 'Mathematics for Engineers' (Wiley, New York, 1949).
- (13) WOODWARD, R. M.: 'Probability and Information Theory' (Pergamon Press, 1953).
- (14) RICE, S. O.: 'Mathematical Analysis of Random Noise', *Bell System Technical Journal*, 1944, 23, p. 282.
- (15) MATHEWS, M. V.: 'A Method of Evaluating Nonlinear Servomechanisms', *Transactions of the American I.E.E.*, 1955, 74, Part II, p. 114.

(11) APPENDICES

(11.1) Tabulation of Particular Integrals

In Reference 3 the following integral is evaluated for various values of n :

$$\mathcal{J}_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{c_{n-1}p^{n-1} \dots c_0}{d_n p^n + \dots d_1 p + d_0} \right|^2 d\omega \quad (28)$$

where $p = j\omega$.

In the paper, particular values are required for $n = 2$ and 4; the following are for the first four integers:

$$\begin{aligned} \mathcal{J}_1 &= \frac{c_0^2}{2d_0d_1} \\ \mathcal{J}_2 &= \frac{c_1^2d_0 + c_0^2d_2}{2d_0d_1d_2} \\ \mathcal{J}_3 &= \frac{c_2^2d_0d_1 + (c_1^2 - 2c_0c_2)d_0d_3 + c_0^2d_2d_3}{2d_0d_3(-d_0d_3 + d_1d_2)} \\ \mathcal{J}_4 &= \frac{c_3^2(-d_0^2d_3 + d_0d_1d_2) + (c_2^2 - 2c_1c_3)d_0d_1d_4 + (c_1^2 - 2c_0c_2)d_0d_3d_4 + c_0^2(-d_1d_4^2 + d_2d_3d_4)}{2d_0d_4(-d_0d_3^2 - d_1^2d_4 + d_1d_2d_3)} \end{aligned}$$

(11.2) Evaluation of Equivalent Gain

The important integral to be evaluated in determining the equivalent gain of eqn. (3) is

$$\int_{-\infty}^{\infty} f(x_i) P_1(x_i) x_i dx_i \quad . \quad . \quad . \quad (29)$$

where $f(x_i)$ is the non-linear characteristic and is assumed known (at least graphically).

For a Gaussian probability distribution the integral becomes

$$\begin{aligned} & \frac{1}{\sigma_s \sqrt{2\pi}} \int_{-\infty}^{\infty} x_i f(x_i) \exp(-x_i^2/2\sigma_s^2) dx_i \\ &= \frac{\sigma_s}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x_i) \exp(-x_i^2/2\sigma_s^2) d(x_i^2/2\sigma_s^2) \\ &= -\frac{\sigma_s}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x_i) d[\exp(-x_i^2/2\sigma_s^2)] \quad . \quad (30) \end{aligned}$$

Integrating by parts,

$$\begin{aligned} & -\frac{\sigma_s}{\sqrt{2\pi}} \left[f(x_i) \exp(-x_i^2/2\sigma_s^2) \right]_{x=-\infty}^{x=\infty} \\ & + \frac{\sigma_s}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-x_i^2/2\sigma_s^2) df(x_i) \end{aligned}$$

The integrated portion is zero when the limits of x_i are applied and hence the integral becomes

$$\frac{\sigma_s}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A(x_i) \exp(-x_i^2/2\sigma_s^2) dx_i \quad . \quad . \quad (31)$$

where $A(x_i) = df(x_i)/dx_i$ and is known.

If $A(x_i)$ is an even function of x_i as occurs in many forms of non-linearities, i.e. $f(x_i)$ is odd, then the integral becomes

$$\frac{2\sigma_s}{\sqrt{2\pi}} \int_0^{\infty} A(x_i) \exp(-x_i^2/2\sigma_s^2) dx_i \quad . \quad . \quad (32)$$

For saturation as shown in Fig. 3 the slope is unity in the range $-V_c < x_i < V_c$ and zero outside.

The integral then becomes $\frac{2\sigma_s}{\sqrt{2\pi}} \int_0^{V_c} \exp(-x_i^2/2\sigma_s^2) dx_i$

$$\text{Hence } k_{eg} = \frac{2}{\sqrt{2\pi}\sigma_s} \int_0^{V_c} \exp(-x_i^2/2\sigma_s^2) dx_i \quad . \quad . \quad (33)$$

(11.3) The Probability Density Function of a Sinusoidal Signal together with Noise

Consider a signal

$$e_s = P_0 \cos \theta \quad . \quad . \quad . \quad (34)$$

Then

$$\delta e_s = -P_0 \sin \theta \delta \theta$$

There are two values of θ giving rise to a particular value e_s in any range:

$$0 < \theta < 2\pi$$

Hence the probability of $P_1(e_s)$ finding the amplitude to be between e_s and $e_s + \delta e_s$ is twice that of the probability $R_1(\theta)$ of θ being between θ and $\theta + \delta \theta$.

Hence

$$P_1(e_s) \delta e_s = 2R_1(\theta) \delta \theta$$

Since it is equally likely for θ to be at any value between 0 and 2π then $R_1(\theta)$ is a constant in this range and zero outside it. Since it must exist in this range,

$$\int_0^{2\pi} R_1(\theta) d\theta = 1$$

$$R_1(\theta) = \frac{1}{2\pi}$$

Therefore, from the above,

$$\begin{aligned} P_1(e_s) \delta e_s &= \frac{1}{\pi} \frac{\delta e_s}{P_0 \sin \theta} \\ P_1(e_s) &= \frac{1}{\pi} \frac{1}{\sqrt{(P_0^2 - e_s^2)}} \text{ for } e_s^2 < P_0^2 \\ &= 0 \text{ for } e_s^2 > P_0^2 \end{aligned} \quad . \quad . \quad (35)$$

Let a random signal $x_i(t)$ have a Gaussian probability distribution

$$q_1(x_i) = \frac{1}{\sigma_N \sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma_N^2}\right) \quad . \quad . \quad (36)$$

Then consider the sum of two signals

$$Y(t) = x_i(t) + e_s \quad . \quad . \quad . \quad (37)$$

where

$$e_s = P_0 \cos \theta = P_0 \cos \omega t$$

The probability distribution of Y , say $Q_1(Y)$, is determined¹⁴ by those of x_i and e_s by a convolution integral

$$Q_1(Y) = \int_{-\infty}^{\infty} P_1(e_s) q_1(Y - e_s) de_s \quad . \quad . \quad (38)$$

The limits of the integral can be reduced to the range $-P_0 < e_s < P_0$ since $P_1(e_s)$ is zero outside this range.

$$\text{Hence } Q_1(Y) = \frac{1}{\sigma_N \pi \sqrt{2\pi}} \int_{-P_0}^{P_0} \frac{\exp[-(x_i - e_s)^2/2\sigma_N^2]}{\sqrt{(P_0^2 - e_s^2)}} de_s$$

Replacing

$$e_s = P_0 \cos \theta$$

$$W_N = \sigma_N^2$$

$$Q_1(Y) = \frac{1}{\pi \sqrt{2\pi W_N}} \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta$$

(11.4) The Equivalent Gain for a Sinusoidal Signal together with Noise

The equivalent gain k_{eg} is given by eqn. (3), for which the first probability distribution for signal and noise is given by eqn. (39). The non-linear function $f(x)$ is defined by eqn. (3) and hence the equivalent gain becomes

$$\begin{aligned} k_{eg} &= \frac{\int_{-\infty}^{\infty} x_i^2 dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta}{\pi \sqrt{2\pi W_N} \int_{-\infty}^{\infty} x_i^2 dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta} \\ &= \frac{\int_{-V_c}^{V_c} x_i dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta}{\pi \sqrt{2\pi W_N} \int_{-V_c}^{V_c} x_i dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta} \\ &+ \frac{1}{\pi \sqrt{2\pi W_N}} \int_{-V_c}^{V_c} x_i^2 dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta \\ &+ \frac{V_c}{\pi \sqrt{2\pi W_N}} \int_{V_c}^{\infty} x_i dx_i \int_0^{\pi} \exp[-(x_i - P_0 \cos \theta)^2/2W_N] d\theta \end{aligned} \quad . \quad . \quad (40)$$

In this equation it is permissible to interchange the order of integration. The inner integrals can then be evaluated by taking μ as the variable of integration, where

$$\mu = \frac{1}{\sqrt{(2W_N)}}(x_i - P_0 \cos \theta) \quad (41)$$

By interchanging the order of integration and by substituting eqn. (41) into eqn. (40) the latter is converted to

$$k_{eg} = \frac{\left[\begin{aligned} & \frac{-V_c}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{-\infty}^{S_1} [\sqrt{(2W_N)}\mu + P_0 \cos \theta] \exp - \mu^2 d\mu \\ & + \frac{1}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{S_1}^{S_2} [\sqrt{(2W_N)}\mu + P_0 \cos \theta]^2 \exp - \mu^2 d\mu \\ & + \frac{V_c}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{S_2}^\infty [\sqrt{(2W_N)}\mu + P_0 \cos \theta] \exp - \mu^2 d\mu \end{aligned} \right]}{\frac{1}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{-\infty}^\infty [\sqrt{(2W_N)}\mu + P_0 \cos \theta]^2 \exp - \mu^2 d\mu} \quad (42)$$

where, for convenience,

$$S_1 = \frac{-V_c - P_0 \cos \theta}{\sqrt{(2W_N)}} \quad S_2 = \frac{V_c - P_0 \cos \theta}{\sqrt{(2W_N)}} \quad (43)$$

Upon integration the value of the denominator is found to be

$$\frac{1}{2}P_0^2 + W_N \quad (44)$$

the mean-square value of the input to the non-linear element.

The integrals in the numerator have no algebraic solution and must be evaluated graphically. This is tedious because of the double integral, and it is convenient to reduce the numerator by evaluating the inner integrals in terms of exponential and error functions.

Proceeding in this manner, the first integral in the numerator is reduced to

$$\begin{aligned} & \frac{-V_c}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{-\infty}^{S_1} [\sqrt{(2W_N)}\mu + P_0 \cos \theta] \exp - \mu^2 d\mu \\ & = \frac{-V_c P_0}{2\pi} \int_0^\pi \cos \theta (1 + \text{sign } S_1 \text{ erf } |S_1|) d\theta \\ & \quad + \frac{V_c}{\pi} \sqrt{\left(\frac{W_N}{2\pi}\right)} \int_0^\pi \exp - S_1^2 d\theta \quad (45) \end{aligned}$$

where in the derivation of this equation use has been made of the function

$$\frac{2}{\sqrt{\pi}} \int_{z_1}^{z_2} \exp - z^2 dz = \text{sign } z^2 \text{ erf } |z^2| - \text{sign } z_2 \text{ erf } |z_2| \quad (46)$$

In precisely the same manner, the third integral in the numerator is reduced to

$$\begin{aligned} & \frac{V_c}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{S_2}^\infty [\sqrt{(2W_N)}\mu + P_0 \cos \theta] \exp - \mu^2 d\mu \\ & = \frac{V_c P_0}{2\pi} \int_0^\pi \cos \theta (1 - \text{sign } S_2 \text{ erf } |S_2|) d\theta \\ & \quad + \frac{V_c}{\pi} \sqrt{\left(\frac{W_N}{2\pi}\right)} \int_0^\pi \exp - S_2^2 d\theta \quad (47) \end{aligned}$$

whilst the last integral to be considered is reduced to

$$\begin{aligned} & \frac{1}{\pi\sqrt{\pi}} \int_0^\pi d\theta \int_{S_1}^{S_2} [\sqrt{(2W_N)}\mu + P_0 \cos \theta]^2 \exp - \mu^2 d\mu \\ & = \frac{P_0^2}{\pi} \int_0^\pi \cos^2 \theta (\text{sign } S_2 \text{ erf } |S_2| - \text{sign } S_1 \text{ erf } |S_1|) d\theta \\ & \quad - \frac{P_0}{\pi} \sqrt{\left(\frac{2W_N}{\pi}\right)} \int_0^\pi \cos \theta (\exp - S_2^2 - \exp - S_1^2) d\theta \\ & \quad - \frac{W_N}{\pi\sqrt{\pi}} \int_0^\pi (S_2 \exp - S_2^2 - S_1 \exp - S_1^2) d\theta \\ & \quad + \frac{W_N}{2\pi} \int_0^\pi (\text{sign } S_2 \text{ erf } |S_2| - \text{sign } S_1 \text{ erf } |S_1|) d\theta \quad (48) \end{aligned}$$

By substituting eqns. (44), (45), (47) and (48) into eqn. (42), the latter becomes

$$k_{eg} = \frac{\left[\begin{aligned} & \frac{P_0^2}{2\pi} \int_0^\pi \cos^2 \theta (\text{sign } S_2 \text{ erf } |S_2| - \text{sign } S_1 \text{ erf } |S_1|) d\theta \\ & - \frac{P_0}{\pi} \sqrt{\left(\frac{2W_N}{\pi}\right)} \int_0^\pi \cos \theta (\exp - S_2^2 - \exp - S_1^2) d\theta \\ & - \frac{W_N}{\pi\sqrt{\pi}} \int_0^\pi (S_2 \exp - S_2^2 - S_1 \exp - S_1^2) d\theta \\ & + \frac{W_N}{2\pi} \int_0^\pi (\text{sign } S_2 \text{ erf } |S_2| - \text{sign } S_1 \text{ erf } |S_1|) d\theta \\ & - \frac{V_c P_0}{2\pi} \int_0^\pi \cos \theta (\text{sign } S_1 \text{ erf } |S_1| + \text{sign } S_2 \text{ erf } |S_2|) d\theta \\ & + \frac{V_c}{\pi} \sqrt{\left(\frac{W_N}{2\pi}\right)} \int_0^\pi (\exp - S_1^2 + \exp - S_2^2) d\theta \end{aligned} \right]}{\frac{1}{2}P_0^2 + W_N} \quad (49)$$

Converting to the normalized form, this expression becomes

$$k_{eg} = \frac{\left[\begin{aligned} & \sqrt{(\pi)} P_{0n}^2 \int_0^\pi \cos \theta (\text{sign } S_{2n} \text{ erf } |S_{2n}| - \text{sign } S_{1n} \text{ erf } |S_{1n}|) d\theta \\ & - 2P_{0n} \sigma_{Nn} \int_0^\pi \cos \theta (\exp - S_{2n}^2 - \exp - S_{1n}^2) d\theta \\ & - \sigma_{Nn}^2 \int_0^\pi (S_{2n} \exp - S_{2n}^2 - S_{1n} \exp - S_{1n}^2) d\theta \\ & + \sqrt{(\pi/2)} \sigma_{Nn}^2 \int_0^\pi (\text{sign } S_{2n} \text{ erf } |S_{2n}| - \text{sign } S_{1n} \text{ erf } |S_{1n}|) d\theta \\ & - \sqrt{(\pi)} P_{0n} \int_0^\pi \cos \theta (\text{sign } S_{2n} \text{ erf } |S_{2n}| + \text{sign } S_{1n} \text{ erf } |S_{1n}|) d\theta \\ & + \sqrt{(2)} \sigma_{Nn} \int_0^\pi (\exp - S_{1n}^2 + \exp - S_{2n}^2) d\theta \end{aligned} \right]}{\pi\sqrt{\pi}(P_{0n}^2 + \sigma_{Nn}^2)} \quad (50)$$

where

$$\sigma_{Nn} = \frac{\sqrt{(2W_N)}}{V_c} \quad . \quad . \quad . \quad (51)$$

and

$$S_{1n} = \frac{-1 - P_{0n} \cos \theta}{\sigma_{Nn}}, S_{2n} = \frac{1 - P_{0n} \cos \theta}{\sigma_{Nn}} \quad . \quad (52)$$

(11.5) The Probability Density Function of a Random Signal contaminated by Noise

Let x_s be the instantaneous signal component.
and x_N be the corresponding noise component.

Let their individual probability density functions be respectively:

$$P_{1s} = \frac{1}{\sigma_s \sqrt{(2\pi)}} \exp(-x_s^2/2\sigma_s^2) \quad . \quad . \quad . \quad (53)$$

and

$$P_{1N} = \frac{1}{\sigma_N \sqrt{(2\pi)}} \exp(-x_N^2/2\sigma_N^2) \quad . \quad . \quad . \quad (54)$$

where σ_s and σ_N are the r.m.s. values of signal and noise respectively.

Let the sum of the two inputs be x_i . The first probability density function is then, from eqn. (38),

$$Q_1(x_i) = \frac{1}{2\pi\sigma_s\sigma_N} \int_{-\infty}^{\infty} \exp\left\{-\left[\frac{r^2}{2\sigma_s^2} + \frac{(x_i - r)^2}{2\sigma_N^2}\right]\right\} dr$$

$$\begin{aligned} &= \frac{1}{2\pi\sigma_s\sigma_N} \int_{-\infty}^{\infty} \exp\left[-\frac{(\sigma_N^2 + \sigma_s^2)}{2\sigma_N^2\sigma_s^2} \left(r^2 + \frac{x_i^2\sigma_s^2 - 2x_i r\sigma_s^2}{\sigma_N^2 + \sigma_s^2}\right)\right] dr \\ &= \frac{1}{2\pi\sigma_s\sigma_N} \int_{-\infty}^{\infty} \exp\left[-\frac{(\sigma_N^2 + \sigma_s^2)}{2\sigma_N^2\sigma_s^2} \left(r - \frac{x_i\sigma_s^2}{\sigma_N^2 + \sigma_s^2}\right)^2 - \frac{x_i^2}{2(\sigma_N^2 + \sigma_s^2)}\right] dr \end{aligned}$$

This equation can be reduced by taking as the new variable

$$z = \frac{\sqrt{(\sigma_s^2 + \sigma_N^2)}}{\sigma_s\sigma_N} \left(r - \frac{x_i\sigma_s^2}{\sigma_N^2 + \sigma_s^2}\right) \quad . \quad . \quad . \quad (55)$$

Hence

$$\begin{aligned} Q_1(x_i) &= \frac{1}{2\pi\sqrt{(\sigma_s^2 + \sigma_N^2)}} \int_{-\infty}^{\infty} \exp\left[-\frac{z^2}{2} - \frac{x_i^2}{2(\sigma_N^2 + \sigma_s^2)}\right] dz \\ &= \frac{1}{\sqrt{(2\pi)}\sqrt{(\sigma_s^2 + \sigma_N^2)}} \exp[-x_i^2/2(\sigma_s^2 + \sigma_N^2)] \quad . \quad (56) \end{aligned}$$

which is the desired result.

THE CHANGE OF AIR-GAP FLUX IN ELECTRICAL MACHINES DUE TO THE DISPLACEMENT OF OPPOSED SLOTS

By G. LIEBMANN, D.Phil., F.Inst.P.

(The paper was first received 2nd August, and in revised form 12th September, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

The change of air-gap flux in electrical machines, due to the relative displacement of rows of opposed slots, is measured, for a series of slot widths, with the aid of the resistance-network analogue. The results are presented in graphical form and can be scaled for different values of slot pitch.

(1) INTRODUCTION

The determination of the magnetic-flux leakage in electrical machines caused by the ventilating ducts of the rotor or the coil slots in rotor and stator is an old problem, which was solved for opposed slots many years ago by F. W. Carter,¹ by the method of conformal transformations. A knowledge of the flux changes due to a relative displacement of slots of various widths and pitches is often desirable, for example, in order to allow an estimate of the ripple waveform in the output of a generator. A more important application (discussed in Dr. Carter's paper), lies in the problem of the mechanical centring stability of the rotor, particularly in those cases where a small oscillation about a stable position is deliberately introduced to reduce commutator wear.

In his paper, Carter showed that the mechanical restoring force is the product of simple geometrical factors and the term $(dB/dx)\Delta B$, where $B = B(x)$ is the average flux density for the slot displacement x , and ΔB is the difference of the flux changes brought about by the presence of the slots, in the extreme positions of zero and very large x . ΔB was calculated rigorously by Carter, but $B(x)$, and its still more important derivative, dB/dx , could only be estimated, since the exact determination of these functions for displaced slots, according to Carter, 'involves elliptic functions in a practically unworkable form'.

An experimental attack on this problem was made a few years ago by L. R. Blake,² who used an electrolytic tank. Although Blake was able to confirm the approximate correctness of Carter's estimated $B(x)$ curve, he found that the accuracy of the electrolytic-tank method was not sufficiently high to obtain reliable data for the dB/dx function, which represents a very small relative effect. As the resistance-network method of field plotting is much more accurate,^{3,4} and the finite mesh arrangement is ideally suited to the representation of models of slotted armatures, the author used this technique to study a number of cases covering the range of practical interest.

(2) DESCRIPTION OF THE METHOD

If the depth of the slots, l , is at least twice their width, d , as is usually the case in practice, any change in l will not affect the flux distribution to a measurable degree unless the iron is highly saturated. The parameter l can therefore be disregarded in this context, provided that the condition $l \geq 2d$ is satisfied; this was

always the case in the present investigation. The problem is then specified by four geometrical parameters (identified in Fig. 1): slot width d , slot pitch t , air-gap g and slot displacement x .

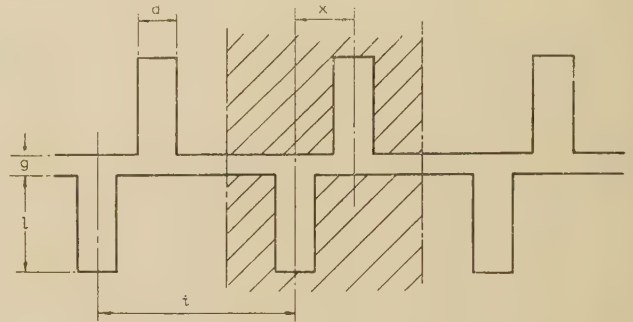


Fig. 1.—Diagram of opposed slots.

$$\frac{d}{g} = \delta \quad \frac{t}{g} = \tau \quad \frac{x}{g} = \xi$$

Uniform values of d and t were assumed, in conformance with normal engineering practice. Non-uniform values of d and t (or values of $l \leq 2d$) could have been investigated by the same technique without any difficulty (subject to certain limitations due to the size of the available analogue equipment), but this would have consumed much time and added little practically useful knowledge. As the flux distribution depends on relative dimensions only, g was chosen as a unit of measurement, so that the relative dimensions of width, pitch and displacement could be written $\delta = d/g$, $\tau = t/g$ and $\xi = x/g$, respectively.

The change of flux in the air-gap, caused by the presence of (uniform) slots spaced at (uniform) distances t , does not depend, for any relative position ξ of the slots, on the relative slot pitch τ , provided that $\tau \geq 2(\delta + 2)$, as is borne out by the data of Figs. 2 and 5. This condition is usually fulfilled in practice. The average flux density B and the change in B due to the slots are most conveniently expressed in terms of the permeance $\Lambda[(= \Lambda(\xi, \delta, \tau)]$ of the air-gap, for an area of unit width and length l , where $\Lambda = \Phi/F$, Φ being the total flux crossing the gap in the specified area, and F the (scalar) magnetic potential difference between the two pole faces.

Once Λ has been determined with sufficient accuracy, both factors in the required term $(dB/dx)\Delta B$ can be evaluated, since $B \propto \Lambda$, $dB/dx \propto d\Lambda/d\xi$ and $\Delta B \propto \Delta\Lambda$, where $\Delta\Lambda = \Lambda_0 - \Lambda_\infty$, and $\Lambda = \Lambda_0$ when $\xi = 0$, $\Lambda = \Lambda_\infty$ when $\xi \geq (\delta + 2)$.

In order to determine Λ for various geometries, models of a representative section of the air-gap region containing one pair of slots, shown cross-hatched in Fig. 1, were set up on a precision resistance network, for different ξ and for different δ . Plots of equal magnetic potential, F , were taken in a few cases, as they give a good idea of the amount of field penetration into the slots, and the nature and range of the disturbance caused by their presence. The measurement of permeance by the resistance-

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Liebmann, who was at the Aldermaston Research Laboratory of Associated Electrical Industries, Limited, died on the 18th June, 1956. The MS. was revised and the proofs were read by G. D. Archard, who is at the same Laboratory.

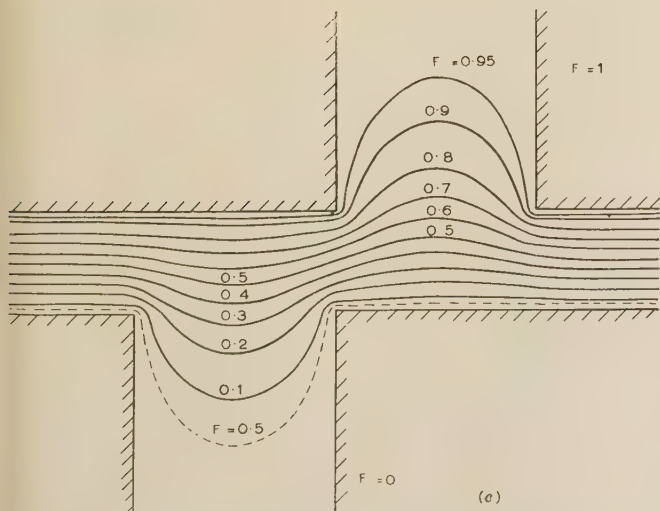
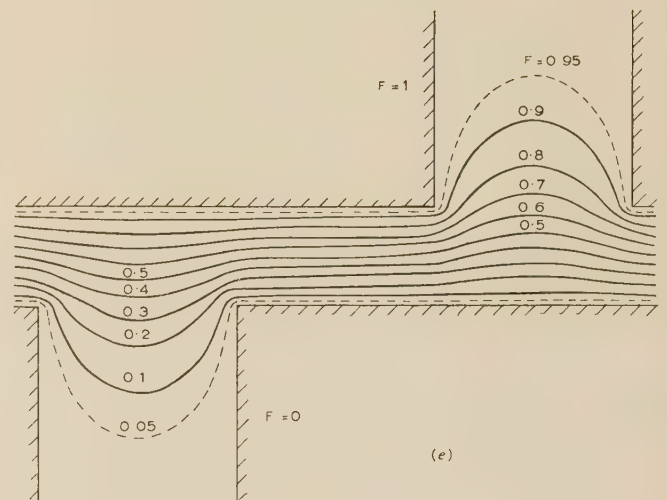
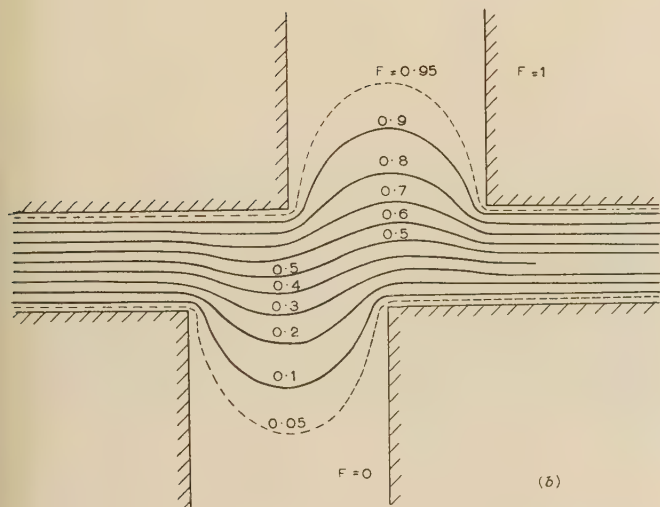
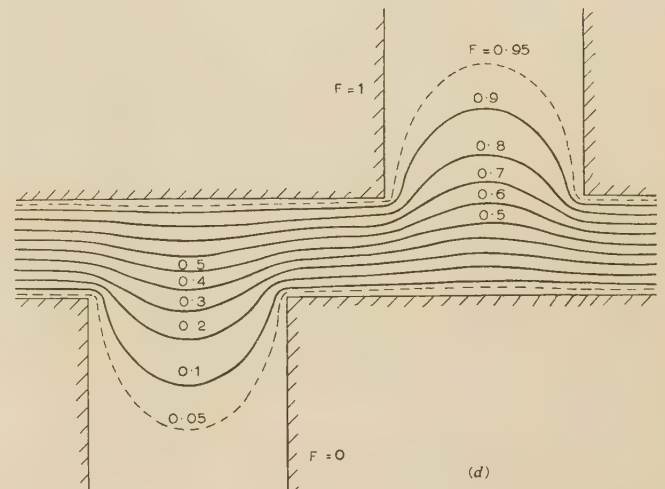
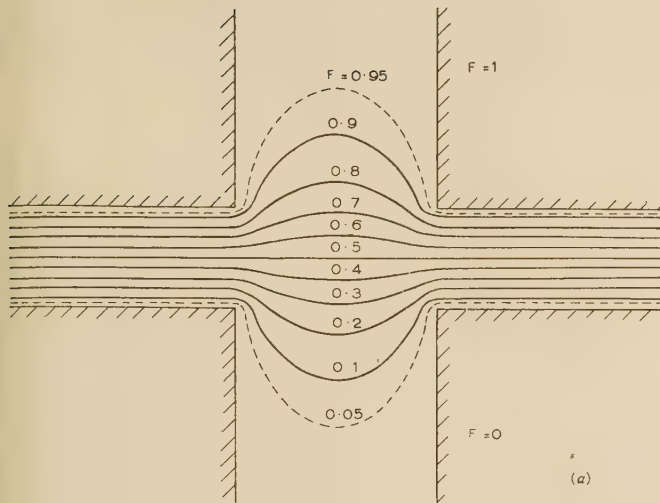


Fig. 2.—Lines of equal magnetic potential for displaced slots.

- (a) $\delta = 2, \delta' = 0$
 (b) $\delta = 2, \delta' = 1$
 (c) $\delta = 2, \delta' = 2$
 (d) $\delta = 2, \delta' = 3$
 (e) $\delta = 2, \delta' = 4$

network method is particularly simple and quick, since $\Lambda \propto 1/R$, so that it is only necessary to take a measurement of the total resistance R between the two conductors outlining the shapes of the rotor and stator sections in the model. An important requirement is that the resistance measurement must be carried out with care and high precision because the chief information sought, namely the change of permeance with relative slot displacement, usually represents a very small effect. The function $d\Lambda/d\xi$ is obtained by numerical differentiation of the measured $1/R$ values. These were determined in each case for two different model sizes, and the previously-described extrapolation technique for the reduction of truncation errors⁴ was applied.

(3) RESULTS

A series of plots of the magnetic-potential distribution across the air-gap and within the slots is given in Fig. 2, the total potential difference F being assumed equal to unity. These plots are typical for the magnetic-field distribution found in slotted air-gaps.

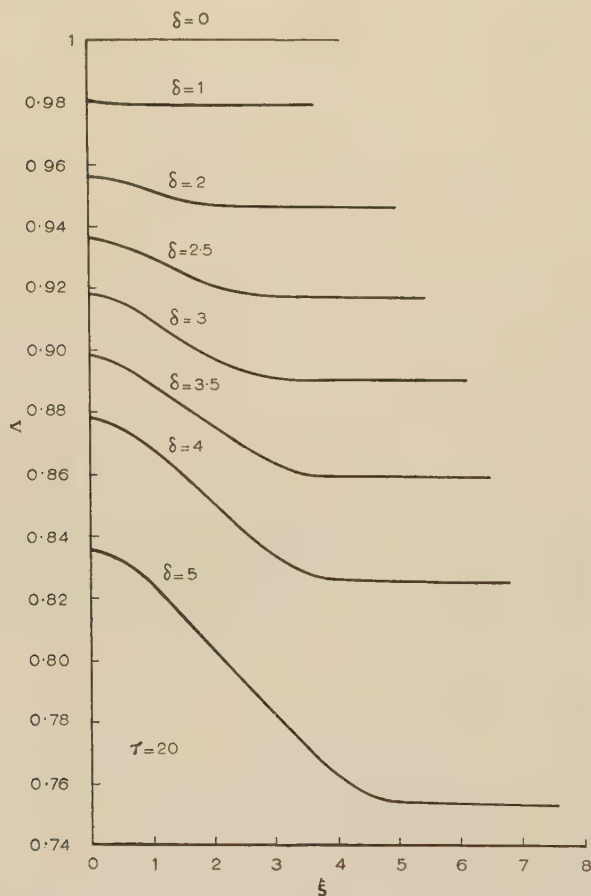


Fig. 3.—Relative permeance Λ as a function of slot displacement ξ for several slot widths δ .

Fig. 3 shows the permeance Λ plotted as a function of displacement ξ for several values of δ , from $\delta = 0$ to $\delta = 5$. These are given in relative units, the permeance for the unslotted air-gap ($\delta = 0$) being taken as $\Lambda = 1$. For all cases, a fixed pitch of $\tau = \tau_0 = 20$ was used. Fig. 3 can be adapted to find the permeance for any other value of τ , subject to the earlier-mentioned restriction $\tau \geq 2(\delta + 2)$, by using the relation

$$\Lambda(\tau) = (1 - \tau_0/\tau)[1 - \Lambda(\tau_0)]$$

where $\Lambda(\tau_0)$ is the value of Λ given in Fig. 3.

$\Delta\Lambda$ can be plotted as a function of δ from the data underlying the curves of Fig. 3. The measured $\Delta\Lambda$ values (for $\tau = \tau_0 = 20$) are shown as circles in Fig. 4. The curve drawn is theoretical

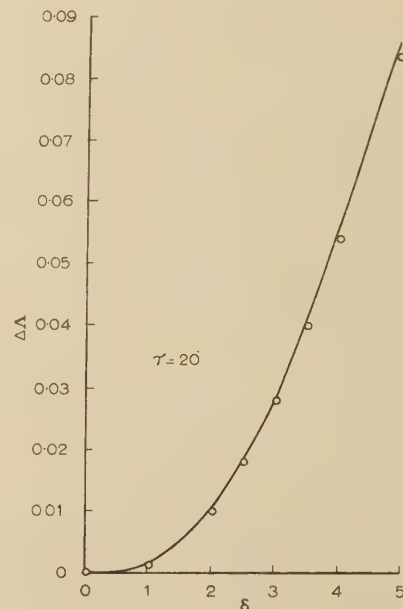


Fig. 4.—Reduction of permeance $\Delta\Lambda$ as a function of slot width δ . Circles denote measured values; continuous line denotes theoretical curve.

and can be evaluated from a modification of Carter's formulae,¹ or from a formula given by Ollendorff.⁵ It is seen that the measured $\Delta\Lambda$ values follow the theoretical curve closely, but a small systematic deviation is noticeable. It is presumed that this is due to the residual truncation error of the representation of the slotted air-gap by a model using finite meshes instead of a true resistive continuum. For $\tau \neq \tau_0$, the graph of Fig. 4 has to be scaled in the ratio $\tau_0 : \tau$, i.e.

$$\Delta\Lambda(\tau) = (\tau_0/\tau)\Delta\Lambda(\tau_0)$$

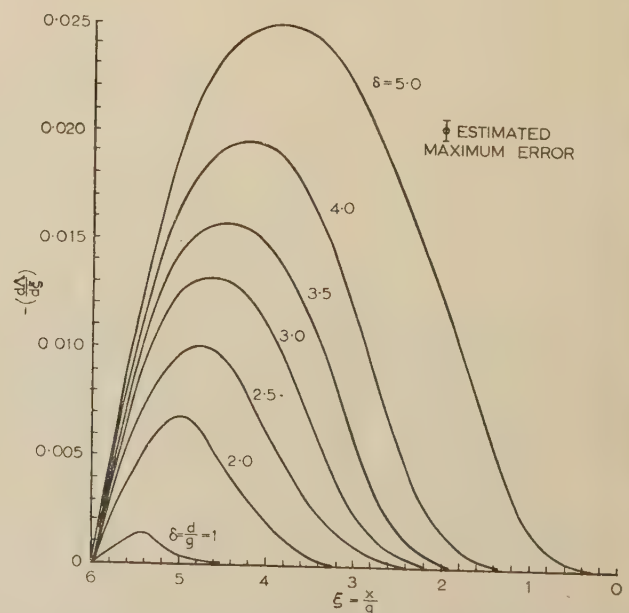


Fig. 5.—Differential change of permeance $-(d\Lambda/d\xi)$ as a function of slot displacement ξ for several slot widths δ ($\tau = 20$).

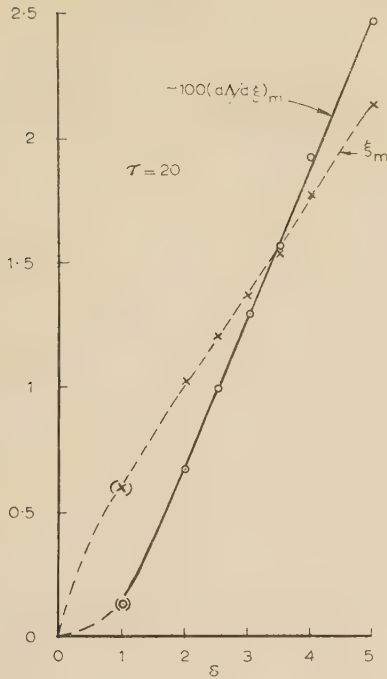


Fig. 6.—Maximum value of $-d\Lambda/d\xi$ and displacement ξ_m for maximum value of $-d\Lambda/d\xi$, as functions of slot width δ .

The further information required to calculate the restoring forces from $(dB/dx)\Delta B$ is contained in Fig. 5, where the factor most sensitive to the slot displacement, $-(d\Lambda/d\xi)$, is plotted as a function of ξ for a range of values of δ . In this Figure, again, $\tau = \tau_0 = 20$. The data of Fig. 5 can be modified for any other value of τ , provided that $\tau \geq 2(\delta + 2)$, by direct scaling in the ratio $\tau_0 : \tau$ since

$$\frac{d\Lambda(\tau)}{d\xi} = \frac{\tau_0}{\tau} \left[\frac{d\Lambda(\tau_0)}{d\xi} \right]$$

Numerical differentiation of measured data increases the relative effect of measuring errors, and the errors in the graphs of Fig. 5 are therefore greater than those in the graphs of Figs. 3 and 4. An estimate of the maximum error in the $d\Lambda/d\xi$ values was made from the estimated errors in the resistance measurements on the analogue models from which the curves of Fig. 5 were derived. This maximum error is shown in graphical form in Fig. 5, the average being considerably smaller. It is noteworthy that Carter's approximate estimate of the general shape of the dB/dx curves (his Fig. 3) agrees very well with the shape of those for $d\Lambda/d\xi$ found in this more detailed quantitative study.

Fig. 6 shows the maximum height $(d\Lambda/d\xi)_m$, and the position ξ_m of the maximum, as a function of δ . It is seen that these relations are linear for $\delta \geq 1.5$, and can therefore be used for convenient interpolation between the curves of Fig. 5.

(4) ACKNOWLEDGMENTS

The author wishes to thank Dr. T. E. Allibone for permission to publish the paper, Mr. A. C. Heath for assistance with the analogue measurements, and Dr. L. R. Blake and Mr. F. A. Haigh of the British Thomson-Houston Company for helpful discussions.

(5) REFERENCES

- (1) CARTER, F. W.: 'Magnetic Centering of Dynamo-Electric Machinery', *Minutes of the Proceedings of The Institution of Civil Engineers*, 1912, **187**, p. 311.
- (2) BLAKE, L. R.: Private communication.
- (3) LIEBMANN, G.: 'Precise Solution of Partial Differential Equations by Resistance Networks', *Nature*, 1949, **164**, p. 149.
- (4) LIEBMANN, G.: 'Solution of Partial Differential Equations with a Resistance Network Analogue', *British Journal of Applied Physics*, 1950, **1**, p. 92.
- (5) OLLENDORFF, F.: 'Potential Felder der Elektrotechnik' (Springer, Berlin, 1932), p. 221, eqn. 71(a).

NON-LINEAR DISTORTION IN TRANSISTOR AMPLIFIERS AT LOW SIGNAL LEVELS AND LOW FREQUENCIES

By N. I. MEYER, Ph.D., M.Sc.

(The paper was first received 10th April, and in revised form 31st July, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

An analysis is made with the purpose of establishing analytical methods of calculating the non-linear distortion of sinusoidal signals in the input and output circuits of transistor amplifiers at low frequencies and low distortion levels. The transistor is regarded as an 'almost linear' four-pole, and characterizing matrix equations and T-network equivalent circuits are developed under the further assumption that the second harmonic is of greater magnitude than the higher harmonics. By using these equations or the equivalent circuits it is possible to calculate second-harmonic distortion of the signals by well-known methods from the linear theory, and it is indicated how the principle may be extended to allow calculation of the third-harmonic distortion also. The principle is used to calculate the output distortion factor for some simple amplifiers, and it is shown that the input and output distortion of the common-emitter amplifier may cancel under special working conditions of the amplifier. From the analytical expressions developed it is possible to fix these conditions so that a considerable reduction of the second harmonic in the output signal is obtained. The expressions developed have been verified experimentally.

LIST OF PRINCIPAL SYMBOLS

F_r = Reduction factor for non-linear distortion.
 i_1 = Amplitude of first-harmonic current.
 i_n = Amplitude of n th-harmonic current.
 R_g = Internal resistance of generator.
 R_l = Load resistance.
 R_p = Parallel feedback resistance.
 R_s = Series feedback resistance.
 v_g = Amplitude of generator voltage.
 v_1 = Amplitude of first-harmonic voltage.
 v_n = Amplitude of n th harmonic voltage.
 r_e, r_b, r_c = T-network parameters.
 a = T-network parameter equal to r_m/r_c .

The upper indices b and e relate the following symbols to the common-base and the common-emitter transistors, respectively.

$h_{11}^b, h_{12}^b, h_{21}^b, h_{22}^b; h_{11}^e, h_{12}^e, h_{21}^e, h_{22}^e$ = h -parameters.
 $K_1^b, K_2^b, K_3^b, K_4^b, K_5^b, K_6^b; K_1^e, K_2^e, K_3^e, K_4^e, K_5^e, K_6^e$ = Distortion constants.
 $z_{11}^b, z_{12}^b, z_{21}^b, z_{22}^b; z_{11}^e, z_{12}^e, z_{21}^e, z_{22}^e$ = z -parameters.
 i_c^b = Alternating collector-current.
 i_{c1}^b, i_{c1}^e = Amplitude of first-harmonic collector-current.
 i_{c2}^b, i_{c2}^e = Amplitude of second-harmonic collector-current.
 i_e^b = Alternating emitter-current.
 i_{e1}^b = Amplitude of first-harmonic emitter-current.
 i_{e2}^b = Amplitude of second-harmonic emitter-current.
 i_{b1}^e = Amplitude of first-harmonic base-current.
 i_{b2}^e = Amplitude of second-harmonic base-current.
 $i_{2\omega}^b, i_{2\omega}^e$ = Second-harmonic current.
 $i_{2\omega a}^b, i_{2\omega a}^e$ = Amplitude of second-harmonic current.
 v_c^b = Alternating collector-voltage.
 v_{c1}^b, v_{c1}^e = Amplitude of first-harmonic collector-voltage.

v_{c2}^b, v_{c2}^e = Amplitude of second-harmonic collector-voltage.
 v_e^b = Alternating emitter-voltage.
 v_{e1}^b = Amplitude of first-harmonic emitter-voltage.
 v_{e2}^b = Amplitude of second-harmonic emitter-voltage.
 v_{b1}^e = Amplitude of first-harmonic base-voltage.
 v_{b2}^e = Amplitude of second-harmonic base-voltage.
 $v_{2\omega}^b, v_{2\omega}^e$ = Second-harmonic voltage.
 $v_{2\omega a}^b, v_{2\omega a}^e$ = Amplitude of second-harmonic voltage.
 $v_{3\omega}^b$ = Third-harmonic voltage.
 A_i^e = Current amplification.
 A_v = Voltage amplification.
 k_d^b = Distortion factor in common-base transistor.
 k_d^e = Distortion factor in common-emitter transistor.
 k_{df}^e = Distortion factor in common-emitter transistor with feedback.

(1) INTRODUCTION

The purpose of the paper has been to develop analytical methods for calculating the non-linear distortion of sinusoidal signals in transistor amplifiers at low frequencies and low signal-levels. It is supposed that the second-harmonic distortion is 5% or less, while third- and higher-harmonic distortion is at least one order of magnitude lower.

It is further assumed that all elements are frequency-independent and real quantities; and that junction transistors only are considered.

In the linear case, the transistor can be either characterized by an equivalent circuit, or regarded as a four-pole and described by some matrix equation. Both possibilities have been applied to the non-linear case treated in the paper, with a view to finding a form on which it is possible to use the well-known calculation methods from linear network theory. This has been achieved, in the case of the equivalent circuit, by including harmonic voltage and current generators in the usual linear circuit and, in the case of the matrix equations, by incorporating some extra terms depending on the square of the first harmonic of some current or voltage in the network.

In Section 2, known principles¹ of calculating the non-linear distortion are applied to a two-pole of one independent variable with the purpose of leading up to the investigation of non-linear four-poles with two independent variables, which case includes the transistor. In earlier quantitative investigations of the non-linear distortion in junction transistors,^{2,3} only a few of the reasons for non-linear distortion are considered, e.g. the input distortion developed by the variation of the input resistance with emitter current or the output distortion developed by the variation of output admittance with collector voltage, while the treatment herein allows a detailed calculation of all the parameters determining the small-signal non-linear distortion at low frequencies.

(2) NON-LINEAR TWO-POLE OF ONE INDEPENDENT VARIABLE

The non-linear two-pole R , as shown in Fig. 1, is characterized by $V = F(I)$ and assumed to be independent of time or other

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Meyer is at the Teleteknisk Forskningslaboratorium (Telecommunication Research Laboratory), Copenhagen.

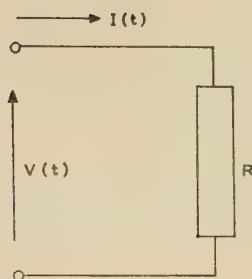


Fig. 1.—Non-linear two-pole.
 $V = F(I)$

parameters. The small-signal impedance corresponding to the working point (V_0, I_0) on such a characteristic is defined as

$$R = \left(\frac{dV}{dI} \right)_{I=I_0}$$

where R will be dependent on I_0 . Only such two-poles as exhibit a slight deviation from linearity, i.e. the so-called 'almost-linear' two-poles¹ will be discussed here, and a stricter definition will be given subsequently.

The voltage and current in Fig. 1 are separated into a time-dependent and a time-independent term thus:

$$V(t) = V_0 + v(t) \quad \text{and} \quad I(t) = I_0 + i(t) \quad (1)$$

Assuming simple characteristics without hysteresis, and given that $i(t) = i_1 \cos \omega t$, we have

$$v(t) = v_0 + v_1 \cos \omega t + v_2 \cos 2\omega t + \dots v_n \cos n\omega t + \dots \quad (2)$$

The condition for the two-pole to be an almost-linear two-pole is given in the expression

$$|v_1| \gg \sum_{n=2}^{\infty} |v_n| \quad (3)$$

This condition can also be expressed in terms of currents if the voltage is supposed to be a sine wave $v(t) = v_1 \cos \omega t$. By analogy with expression (3) the condition then reads

$$|i_1| \gg \sum_{n=2}^{\infty} |i_n| \quad (4)$$

By expanding $V = F(I)$ in a Taylor's series, another expression for v is obtained:

$$V - V_0 = v = \frac{dF}{dI} i + \frac{1}{2!} \frac{d^2F}{dI^2} i^2 + \dots = \sum_{n=1}^{\infty} R_n i^n \quad (5)$$

where $R_n = \frac{1}{n!} \frac{d^n F}{dI^n}$

and the derivatives are to be taken at the point $I = I_0$.

The condition in expression (3) can be expressed in terms of the Taylor coefficients when the current is supposed to be sinusoidal $i = i_1 \cos \omega t$; thus

$$|R_1 i_1| \gg \sum_{n=2}^{\infty} |R_n i_1^n| \quad (6)$$

From expression (6) it is seen that the condition equation for an almost-linear two-pole concerns not only the characteristic of the two-pole but also the signal level.

When the almost-linear two-pole R is connected to a sine-wave voltage generator with linear internal resistance R_g as shown

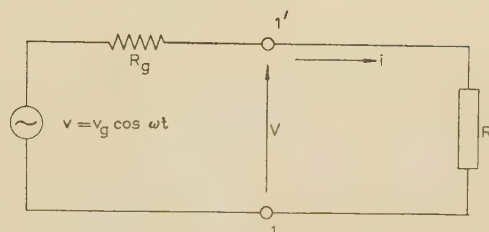


Fig. 2.—Non-linear two-pole connected to sinusoidal voltage generator with linear internal resistance.

in Fig. 2, neither v nor i will be sinusoidal, and it will now be shown how the harmonics can be calculated by taking into consideration the conditions given above.

The quantities v and i can be expressed thus:

$$v = \sum_{n=0}^{\infty} v_n \cos n\omega t$$

and

$$i = \sum_{n=0}^{\infty} i_n \cos n\omega t \quad (7)$$

and, from eqn. (5),

$$v = \sum_{p=1}^{\infty} R_p i^p = \sum_{p=1}^{\infty} R_p \left(\sum_{n=0}^{\infty} i_n \cos n\omega t \right)^p \quad (8)$$

As can be seen from expressions (3), (4) and (6), the terms in eqn. (8) will be of a range of orders of magnitude. Dropping the terms of orders higher than the first two, we get

First-order terms:

$$R_1 i_1 \cos \omega t$$

Second-order terms:

$$R_1 \sum_{n=0}^{\infty} i_n \cos n\omega t + \sum_{p=2}^{\infty} R_p (i_1 \cos \omega t)^p$$

$$i.e. \quad v \simeq R_1 \sum_{n=0}^{\infty} i_n \cos n\omega t + \sum_{p=2}^{\infty} R_p (i_1 \cos \omega t)^p \quad (9)$$

The first term of eqn. (9) corresponds to the voltage developed across 1-1' in Fig. 2 when the current including the harmonics flows through the linear resistance R_1 corresponding to the small-signal impedance of the almost-linear two-pole at the working point $I = I_0$. The second term of eqn. (9) represents a number of harmonic voltages across 1-1', and comparison with eqn. (5) shows that these are the harmonic voltages developed when a sine-wave current $i = i_1 \cos \omega t$ flows through the almost-linear two-pole.

It follows from the above discussion that the almost-linear two-pole can be represented by a linear resistance R_1 in series with a voltage generator of electromotive force given by

$$v_s = \sum_{p=2}^{\infty} R_p (i_1 \cos \omega t)^p \quad (10)$$

as shown in Fig. 3.

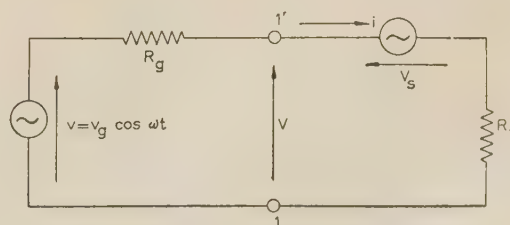


Fig. 3.—Equivalent circuit of the non-linear two-pole.

The currents and voltages of the different harmonics can now be calculated from Fig. 3 by conventional network methods known from the linear theory.

The following Sections, where the principle is applied to transistor circuits, deal mainly with the cases where the second harmonic is of much greater magnitude than the higher harmonics. As an example the currents and voltages in Fig. 3 will be calculated under this assumption, which can be expressed as an extension of the conditions defined by expressions (3), (4) and (6); thus

$$\begin{aligned} |i_1| &\gg |i_2| \gg \sum_{n=1,2}^{\infty} |i_n|; \quad |v_1| \gg |v_2| \gg \sum_{n=0}^{\infty} |v_n|; \\ |R_1 i_1| &\gg |R_2 i_1^2| \gg \sum_{n=3}^{\infty} |R_n i_1^n| \quad \dots \quad (11) \end{aligned}$$

From eqn. (10),

$$v_s \approx \frac{1}{2} R_2 i_1^2 \cos 2\omega t = \frac{1}{4} \frac{d^2 F}{dI^2} i_1^2 \cos 2\omega t$$

and from eqn. (9),

$$v \approx \frac{dF}{dI} (i_1 \cos \omega t + i_2 \cos 2\omega t) + \frac{1}{4} \frac{d^2 F}{dI^2} i_1^2 \cos 2\omega t \quad (12)$$

It is seen that the expressions for the first and second harmonics of the voltage v are obtained by taking into account the first and second terms in the Taylor's-series representation of the almost-linear two-pole. Furthermore, the expression for the current must include the second harmonic in the first-order term, while it is only necessary to use the first harmonic of the current in the second-order term.

When v_g , R_g , R_1 and R_2 are known quantities, it is possible first to calculate i_1 from

$$i_1 = \frac{v_g}{R_g + R_1} \quad \dots \quad (13)$$

and then i_2 thus:

$$i_2 = \frac{v_s}{R_g + R_1} = \frac{1}{2} \frac{R_2 v_g^2}{(R_g + R_1)^3} \quad \dots \quad (14)$$

It is to be remembered that R_2 is measured in ohms per ampere.

It should be mentioned that the example may be extended to include the case in which the magnitude of the third harmonic is the same as, or higher than, that of the second harmonic. In this case expressions (11) may be rewritten

$$\begin{aligned} |i_1| &\gg |i_2|, \quad |i_3| \gg \sum_{n=0}^{\infty} |i_n|; \quad |v_1| \gg |v_2|, \quad |v_3| \gg \sum_{n=0}^{\infty} |v_n|; \\ |R_1 i_1| &\gg |R_2 i_1^2|, \quad |R_3 i_1^3| \gg \sum_{n=4}^{\infty} |R_n i_1^n| \quad \dots \quad (15) \end{aligned}$$

while eqn. (12) should be

$$\begin{aligned} v &\approx \frac{dF}{dI} (i_1 \cos \omega t + i_2 \cos 2\omega t + i_3 \cos 3\omega t) \\ &\quad + \frac{1}{2!} \frac{d^2 F}{dI^2} (i_1 \cos \omega t)^2 + \frac{1}{3!} \frac{d^3 F}{dI^3} (i_1 \cos \omega t)^3 \\ &\approx \frac{dF}{dI} (i_1 \cos \omega t + i_2 \cos 2\omega t + i_3 \cos 3\omega t) \\ &\quad + \frac{1}{4} \frac{d^2 F}{dI^2} i_1^2 \cos 2\omega t + \frac{1}{24} \frac{d^3 F}{dI^3} i_1^3 \cos 3\omega t \quad \dots \quad (16) \end{aligned}$$

(3) THE TRANSISTOR REGARDED AS AN ALMOST-LINEAR FOUR-POLE

The principles developed for an almost-linear two-pole with one independent variable can be extended to an almost-linear four-pole with two independent variables. The definition of an almost-linear four-pole is framed by analogy with that stated in Section 2, and the necessary calculations will follow the same lines. The general case may be ignored for the present and a specific example may be studied, namely the common-base transistor shown in Fig. 4. It is assumed that the signal level is

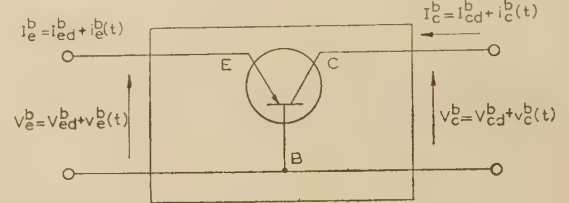


Fig. 4.—The common-base transistor regarded as a four-pole.

$$V_e^b = F(I_e^b, V_c^b) \text{ and } I_c^b = G(I_e^b, V_c^b)$$

so low that the second-harmonic distortion is less than 5%, while distortions due to the higher harmonics are at least one order of magnitude lower. The case in which the third harmonic is of significance will be mentioned at the end of this Section.

The upper index b relates to common-base configuration.

In the linear four-pole theory, the choice of the independent variables I_e^b and V_c^b would correspond to characterizing the transistor by the so-called h -matrix (see Reference 2, page 304):

$$\begin{bmatrix} v_e^b \\ i_c^b \end{bmatrix} = \begin{bmatrix} \frac{\partial F}{\partial I_e^b} & \frac{\partial F}{\partial V_c^b} \\ \frac{\partial G}{\partial I_e^b} & \frac{\partial G}{\partial V_c^b} \end{bmatrix} \begin{bmatrix} i_e^b \\ v_c^b \end{bmatrix} = \begin{bmatrix} h_{11}^b & h_{12}^b \\ h_{21}^b & h_{22}^b \end{bmatrix} \begin{bmatrix} i_e^b \\ v_c^b \end{bmatrix} \quad (17)$$

where alternating voltages and currents are denoted by small letters.

When the second harmonics are to be calculated the same principles as indicated in Section 2 in connection with eqn (12) can be used, i.e. the Taylor's series is expanded to include the second-order terms:

$$v_e^b = \frac{\partial F}{\partial I_e^b} i_e^b + \frac{\partial F}{\partial V_c^b} v_c^b + \frac{1}{2} \frac{\partial^2 F}{\partial I_e^b \partial I_e^b} (i_e^b)^2 + \frac{1}{2} \frac{\partial^2 F}{\partial V_c^b \partial V_c^b} (v_c^b)^2 + \frac{\partial^2 F}{\partial I_e^b \partial V_c^b} i_e^b v_c^b \quad \dots \quad (18)$$

$$i_c^b = \frac{\partial G}{\partial I_e^b} i_e^b + \frac{\partial G}{\partial V_c^b} v_c^b + \frac{1}{2} \frac{\partial^2 G}{\partial I_e^b \partial I_e^b} (i_e^b)^2 + \frac{1}{2} \frac{\partial^2 G}{\partial V_c^b \partial V_c^b} (v_c^b)^2 + \frac{\partial^2 G}{\partial I_e^b \partial V_c^b} i_e^b v_c^b \quad \dots \quad (19)$$

where

$$\left. \begin{aligned} v_e^b &= v_{e1}^b \cos \omega t + v_{e2}^b \cos 2\omega t \\ i_e^b &= i_{e1}^b \cos \omega t + i_{e2}^b \cos 2\omega t \\ v_c^b &= v_{c1}^b \cos \omega t + v_{c2}^b \cos 2\omega t \\ i_c^b &= i_{c1}^b \cos \omega t + i_{c2}^b \cos 2\omega t \end{aligned} \right\} \quad \dots \quad (20)$$

By analogy with the procedure in Section 2, the approximate expressions of the dependent variables are found by inserting the expressions including the first and second harmonics of the independent variables into the first-order terms, and those including the first harmonic of the independent variables into the second-order terms thus:

$$\left. \begin{aligned} v_e^b &= h_{11}^b i_{e1}^b + h_{12}^b v_{c1}^b + v_{e2\omega}^b \\ i_c^b &= h_{21}^b i_{e1}^b + h_{22}^b v_{c1}^b + i_{c2\omega}^b \end{aligned} \right\} \quad \dots \quad (21)$$

where

$$i_{e2\omega}^b = \left[\frac{1}{4} \frac{\partial h_{11}^b}{\partial I_e^b} (i_{e1}^b)^2 + \frac{1}{4} \frac{\partial h_{12}^b}{\partial V_c^b} (v_{c1}^b)^2 + \frac{1}{2} \frac{\partial h_{11}^b}{\partial V_c^b} i_{e1}^b v_{c1}^b \right] \cos 2\omega t$$

$$\equiv [K_1^b (i_{e1}^b)^2 + K_2^b i_{e1}^b v_{c1}^b + K_3^b (v_{c1}^b)^2] \cos 2\omega t \quad (22)$$

and

$$i_{e2\omega}^b = \left[\frac{1}{4} \frac{\partial h_{21}^b}{\partial I_e^b} (i_{e1}^b)^2 + \frac{1}{4} \frac{\partial h_{22}^b}{\partial V_c^b} (v_{c1}^b)^2 + \frac{1}{2} \frac{\partial h_{21}^b}{\partial V_c^b} i_{e1}^b v_{c1}^b \right] \cos 2\omega t$$

$$\equiv [K_4^b (i_{e1}^b)^2 + K_5^b i_{e1}^b v_{c1}^b + K_6^b (v_{c1}^b)^2] \cos 2\omega t \quad (23)$$

From linear four-pole theory expressions are available relating the first harmonics of the currents and voltages for the terminated four-pole, and if i_{e1}^b and v_{c1}^b are expressed in terms of i_{c1}^b , we have

$$i_{e1}^b = \frac{1 + h_{22}^b R_l}{h_{21}^b} i_{c1}^b \text{ and } v_{c1}^b = -R_l i_{c1}^b \quad (24)$$

where R_l is the load resistance (see Fig. 5). Substitution in eqns. (22) and (23) yields

$$\left. \begin{aligned} v_{2\omega}^b &= v_{2\omega a}^b \cos 2\omega t = K_A (i_{c1}^b)^2 \cos 2\omega t \\ i_{2\omega}^b &= i_{2\omega a}^b \cos 2\omega t = K_B (i_{c1}^b)^2 \cos 2\omega t \end{aligned} \right\} \quad (25)$$

$$\text{where } K_A = K_1^b \frac{(1 + h_{22}^b R_l)^2}{(h_{21}^b)^2} - K_2^b \frac{(1 + h_{22}^b R_l)}{h_{21}^b} R_l + K_3^b R_l^2 \quad (26)$$

$$\text{and } K_B = K_4^b \frac{(1 + h_{22}^b R_l)^2}{(h_{21}^b)^2} - K_5^b \frac{(1 + h_{22}^b R_l)}{h_{21}^b} R_l + K_6^b R_l^2 \quad (27)$$

Using eqns. (21) and (25) the characterizing matrix equations can be written, and the first and second harmonics of the currents and voltages can be calculated from them by well-known matrix methods thus:

$$\begin{bmatrix} v_{e1}^b \\ i_{c1}^b \end{bmatrix} = \begin{bmatrix} h_{11}^b & h_{12}^b \\ h_{21}^b & h_{22}^b \end{bmatrix} \begin{bmatrix} i_{e1}^b \\ v_{c1}^b \end{bmatrix} \quad (28)$$

$$\begin{bmatrix} v_{e2}^b - v_{2\omega a}^b \\ i_{c2}^b - i_{2\omega a}^b \end{bmatrix} = \begin{bmatrix} h_{11}^b & h_{12}^b \\ h_{21}^b & h_{22}^b \end{bmatrix} \begin{bmatrix} i_{e2}^b \\ v_{c2}^b \end{bmatrix} \quad (29)$$

The use of these matrix equations is illustrated by calculating the distortion factor of the output current in Fig. 5. To eliminate i_{e2}^b and v_{c2}^b , we have

$$\left. \begin{aligned} v_{e2}^b &= -R_g i_{e2}^b \\ v_{c2}^b &= -R_l i_{c2}^b \end{aligned} \right\} \quad (30)$$

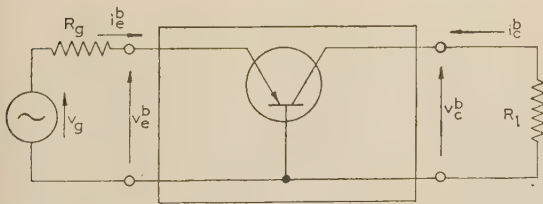


Fig. 5.—Terminated transistor four-pole.

By inserting eqns. (25) and (30) in eqn. (29) and performing simple matrix manipulations, the following equation is obtained:

$$\begin{bmatrix} K_A (i_{c1}^b)^2 \\ K_B (i_{c1}^b)^2 \end{bmatrix} = \begin{bmatrix} -(R_g + h_{11}^b) & R_l h_{12}^b \\ -h_{21}^b & 1 + R_l h_{22}^b \end{bmatrix} \begin{bmatrix} i_{e2}^b \\ v_{c2}^b \end{bmatrix} \quad (31)$$

Solving with respect to i_{e2}^b ,

$$i_{e2}^b = \frac{-K_B (R_g + h_{11}^b) + h_{21}^b K_A (i_{c1}^b)^2}{D_1^b} \quad (32)$$

where

$$D_1^b = -[R_l \Delta_h^b + R_g (1 + h_{22}^b R_l) + h_{11}^b]$$

and

$$\Delta_h^b = h_{11}^b h_{22}^b - h_{12}^b h_{21}^b$$

From eqn. (32), the expression for the distortion factor is found to be

$$k_d^b = \frac{|i_{e2}^b|}{|i_{c1}^b|} = \left| \frac{h_{21}^b K_A - (R_g + h_{11}^b) K_B}{D_1^b} \right| i_{c1}^b \quad (33)$$

where K_A and K_B are known quantities dependent on the transistor and the load resistance as shown in eqns. (26) and (27).

In the case of the common-emitter configuration shown in Fig. 6 the whole procedure can be repeated, the only difference

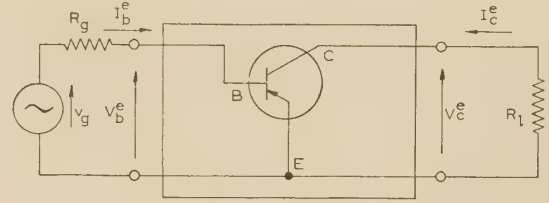


Fig. 6.—Common-emitter amplifier.

being a change of the upper indices from b to e and of some of the lower indices from e to b . The characterizing matrix equations can be written down immediately by comparison with eqns. (28) and (29) thus:

$$V_b^e = H(I_b^e, V_c^e) \text{ and } I_c^e = K(I_b^e, V_c^e)$$

$$\begin{bmatrix} v_{b1}^e \\ i_{c1}^e \end{bmatrix} = \begin{bmatrix} h_{11}^e & h_{12}^e \\ h_{21}^e & h_{22}^e \end{bmatrix} \begin{bmatrix} i_{b1}^e \\ v_{c1}^e \end{bmatrix} \quad (34)$$

$$\begin{bmatrix} v_{b2}^e - v_{2\omega a}^e \\ i_{c2}^e - i_{2\omega a}^e \end{bmatrix} = \begin{bmatrix} h_{11}^e & h_{12}^e \\ h_{21}^e & h_{22}^e \end{bmatrix} \begin{bmatrix} i_{b2}^e \\ v_{c2}^e \end{bmatrix} \quad (35)$$

where

$$v_{2\omega a}^e = \frac{1}{4} \frac{\partial h_{11}^e}{\partial I_b^e} (i_{b1}^e)^2 + \frac{1}{2} \frac{\partial h_{12}^e}{\partial V_c^e} i_{b1}^e v_{c1}^e + \frac{1}{4} \frac{\partial h_{11}^e}{\partial V_c^e} (v_{c1}^e)^2$$

$$\equiv K_1^e (i_{b1}^e)^2 + K_2^e i_{b1}^e v_{c1}^e + K_3^e (v_{c1}^e)^2 \quad (36)$$

$$i_{2\omega a}^e = \frac{1}{4} \frac{\partial h_{21}^e}{\partial I_b^e} (i_{b1}^e)^2 + \frac{1}{2} \frac{\partial h_{22}^e}{\partial V_c^e} i_{b1}^e v_{c1}^e + \frac{1}{4} \frac{\partial h_{21}^e}{\partial V_c^e} (v_{c1}^e)^2$$

$$\equiv K_4^e (i_{b1}^e)^2 + K_5^e i_{b1}^e v_{c1}^e + K_6^e (v_{c1}^e)^2 \quad (37)$$

By analogy with eqn. (33) the distortion factor of the output current in Fig. 6 becomes

$$k_d^e = \frac{|i_{c2}^e|}{|i_{c1}^e|} = \left| \frac{h_{21}^e K_c - (R_g + h_{11}^e) K_D}{D_1^e} \right| \quad (38)$$

where

$$D_1^e = -[R_l \Delta_h^e + R_g (1 + h_{22}^e R_l) + h_{11}^e]$$

$$K_c = K_1^e \left(\frac{1 + h_{22}^e R_l}{h_{21}^e} \right)^2 - K_2^e \frac{1 + h_{22}^e R_l}{h_{21}^e} R_l + K_3^e R_l^2$$

$$\text{and } K_D = K_4^e \left(\frac{1 + h_{22}^e R_l}{h_{21}^e} \right)^2 - K_5^e \frac{1 + h_{22}^e R_l}{h_{21}^e} R_l + K_6^e R_l^2$$

It is possible to use one of the sets $(v_{2\omega a}^b, i_{2\omega a}^b)$ or $(v_{2\omega a}^e, i_{2\omega a}^e)$ in the equations for both the common-base and the common-emitter configurations. Using the set $(v_{2\omega a}^b, i_{2\omega a}^b)$, eqn. (34) will remain unchanged, while the new form of eqn. (35) is found by making the following substitutions in eqn. (29):

$$i_c^e = i_c^b; \quad v_e^e = -v_b^e; \quad v_c^e = v_c^e - v_b^e; \quad i_b^e = -(i_b^e + i_c^e) \quad (39)$$

Using the relation between $\{h^b\}$ and $\{h^e\}$,

$$\begin{bmatrix} h_{11}^e & h_{12}^e \\ h_{21}^e & h_{22}^e \end{bmatrix} = \frac{1}{D_b} \begin{bmatrix} h_{11}^b & \Delta_h^b - h_{12}^b \\ -\Delta_h^b - h_{21}^b & h_{22}^b \end{bmatrix} \quad (40)$$

where $D_b = 1 + \Delta_h^b + h_{21}^b - h_{12}^b \approx 1 + h_{21}^b$

we get, from eqns. (39) and (29),

$$\begin{bmatrix} v_{b2}^e - [h_{11}^e i_{2\omega a}^b - (1 - h_{12}^e) v_{2\omega a}^b] \\ i_{c2}^e - [h_{22}^e v_{2\omega a}^b + (1 + h_{21}^e) i_{2\omega a}^b] \end{bmatrix} = \begin{bmatrix} h_{11}^e & h_{12}^e \\ h_{21}^e & h_{22}^e \end{bmatrix} \begin{bmatrix} i_{b2}^e \\ v_{c2}^e \end{bmatrix} \quad (41)$$

where $1 - h_{12}^e$ for all practical purposes is equal to 1. From eqns. (35) and (41) we get

$$v_{2\omega a}^e = h_{11}^e i_{2\omega a}^b - (1 - h_{12}^e) v_{2\omega a}^b \quad (42)$$

$$\text{and } i_{2\omega a}^e = h_{22}^e v_{2\omega a}^b + (1 + h_{21}^e) i_{2\omega a}^b$$

From eqns. (25), (38) and (42), the distortion factor of the output current in Fig. 6 can be determined in terms of K_A and K_B thus:

$$k_d^e = \left| \frac{i_{c2}^e}{i_{c1}^e} \right| = \left| \frac{h_{21}^e [h_{11}^e K_B - (1 - h_{12}^e) K_A] - (R_g + h_{11}^e) [h_{22}^e K_A + (1 + h_{21}^e) K_B]}{D_2} \right| i_{c1}^e$$

$$\approx \left| \frac{h_{21}^e K_A + K_B [R_g(1 + h_{21}^e) + h_{11}^e]}{D_2} \right| i_{c1}^e \quad (43)$$

where $D_2 = [R_g \Delta_h^e + R_g(1 + h_{22}^e R_f) + h_{11}^e]$

and $h_{12}^e \ll 1$; $h_{21}^e \gg (R_g + h_{11}^e) h_{22}^e$

It should be mentioned that the principle can be extended to include the third harmonic in the same way as indicated at the end of Section 2. This would require introduction of the extra terms

$$v_e' = \frac{1}{6} \frac{\partial^3 F}{\partial I_e^3} (i_e^b)^3 + \frac{1}{2} \frac{\partial^3 F}{\partial I_e^2 \partial V_c} (i_e^b)^2 v_c^b + \frac{1}{2} \frac{\partial^3 F}{\partial I_e \partial V_c^2} i_e^b (v_c^b)^2 + \frac{1}{6} \frac{\partial^3 F}{\partial V_c^3} (v_c^b)^3$$

in eqn. (18), and similarly in eqn. (19).

In the expression for v_e^b in eqn. (21), the following extra terms would have to be inserted:

$$v_{3\omega}^b = K_{13}^b (i_{e1}^b)^3 + K_{23}^b (i_{e1}^b)^2 v_{c1}^b + K_{33}^b i_{e1}^b (v_{c1}^b)^2 + K_{43}^b (v_{c1}^b)^3$$

Analogous terms would have to be introduced in the expressions for i_c^b ; in this way eight new distortion constants would appear.

All calculations concerning the third-harmonic distortion may then be carried out by analogy with the second-harmonic distortion calculations.

(4) THE T-NETWORK EQUIVALENT CIRCUIT INCLUDING SECOND-HARMONIC GENERATORS

In this Section an equivalent circuit is developed which resembles the usual T-network, except that it includes some second-harmonic voltage (or current) generators. To begin with, eqns. (21) are written in matrix form thus:

$$\begin{bmatrix} v_e^b - v_{2\omega}^b \\ i_c^b - i_{2\omega}^b \end{bmatrix} = \begin{bmatrix} h_{11}^b & h_{12}^b \\ h_{21}^b & h_{22}^b \end{bmatrix} \begin{bmatrix} i_e^b \\ v_c^b \end{bmatrix} \quad (44)$$

If this equation is solved with respect to $v_e^b - v_{2\omega}^b$ and v_c^b , we get

$$\begin{bmatrix} v_e^b - v_{2\omega}^b \\ v_c^b \end{bmatrix} = \begin{bmatrix} z_{11}^b & z_{12}^b \\ z_{21}^b & z_{22}^b \end{bmatrix} \begin{bmatrix} i_e^b \\ i_c^b - i_{2\omega}^b \end{bmatrix} \quad (45)$$

where the relation between $\{h^b\}$ and $\{z^b\}$ is given in available matrix tables (see, e.g., Reference 2, page 335). By means of simple matrix manipulations eqn. (45) can be written in the form

$$\begin{bmatrix} v_e^b - v_{2\omega}^b + z_{12}^b i_{2\omega}^b \\ v_c^b + z_{22}^b i_{2\omega}^b \end{bmatrix} = \begin{bmatrix} z_{11}^b & z_{12}^b \\ z_{21}^b & z_{22}^b \end{bmatrix} \begin{bmatrix} i_e^b \\ i_c^b \end{bmatrix} \quad (46)$$

From linear theory we know that the following equation corresponds to the T-network shown in Fig. 7:

$$\begin{bmatrix} v_e^b \\ v_c^b \end{bmatrix} = \begin{bmatrix} z_{11}^b & z_{12}^b \\ z_{21}^b & z_{22}^b \end{bmatrix} \begin{bmatrix} i_e^b \\ i_c^b \end{bmatrix} \quad (47)$$

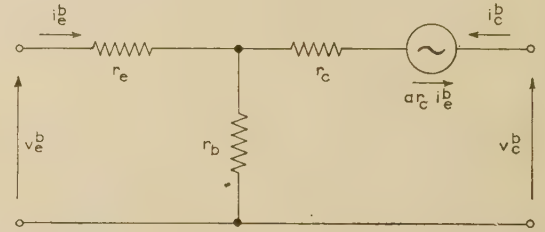


Fig. 7.—Small-signal T-network equivalent circuit of the common-base configuration.

Comparison of eqns. (46) and (47) shows that the T-network corresponding to eqn. (46) will be as shown in Fig. 8, where the following relations² between T-circuit and z parameters have been used:

$$z_{12}^b = r_b \text{ and } z_{22}^b = r_b + r_c$$

The terms $v_{2\omega}^b$ and $i_{2\omega}^b$ are known quantities under specified working conditions for the transistor and can be determined from eqns. (22) and (23).

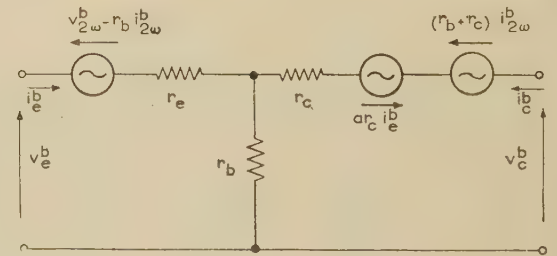


Fig. 8.—T-network equivalent circuit of the common-base configuration with second-harmonic voltage generators.

From Fig. 8 it is now possible to calculate the contents of first and second harmonics in the output and input circuits when the transistor is fed from a voltage generator of arbitrary linear internal resistance and terminated with an arbitrary linear load resistance.

The voltage generators may be inserted in the T-network of Fig. 8 in several different ways, or they may be replaced by current generators. However, let it suffice to show one other example involving a current generator instead of a voltage generator in the output network; this is illustrated in Fig. 9, in which $(r_b + r_c)/r_c \approx 1$ for practical junction transistors.

In the common-emitter case, the T-network equivalent circuit can be developed analogously. Corresponding to eqn. (46) we now get

$$\begin{bmatrix} v_e^e - v_{2\omega}^e + z_{12}^e i_{2\omega}^e \\ v_c^e + z_{22}^e i_{2\omega}^e \end{bmatrix} = \begin{bmatrix} z_{11}^e & z_{12}^e \\ z_{21}^e & z_{22}^e \end{bmatrix} \begin{bmatrix} i_e^e \\ i_c^e \end{bmatrix}$$

From this it will be found that the usual common-emitter equivalent circuit shown in Fig. 10 may be replaced by the

the mesh equations for the second-harmonic currents and voltages of Fig. 14 will yield

$$i_{c2}^e = \frac{v_{2\omega}^b [R_p(ar_c + R_g + r_b) + R_g(r_b + r_c)] + R_p r_c i_{2\omega}^b [r_e + R_s + r_b(1 - a + \frac{r_e + R_s}{r_c}) + R_g]}{D} \quad (48)$$

circuit shown in Fig. 11, where the following relations between T-circuit and z-parameters have been used:

$$\begin{aligned} z_{12}^e &= r_e \\ z_{22}^e &= r_e + r_c(1 - a) \end{aligned}$$

$$\begin{aligned} D &= (r_e + R_s)[(r_b + r_c)(R_p + R_g + R_l) + R_p(R_g + R_l)] \\ &+ r_c\{(1 - a)[r_b(R_p + R_l + R_g) + R_g R_p] + R_l R_g\} \\ &+ R_l[R_p r_b + R_g r_b + R_g R_p] \end{aligned}$$

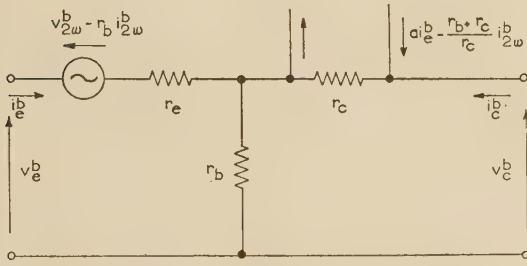


Fig. 9.—T-network equivalent circuit of the common-base configuration with second-harmonic voltage and current generators.

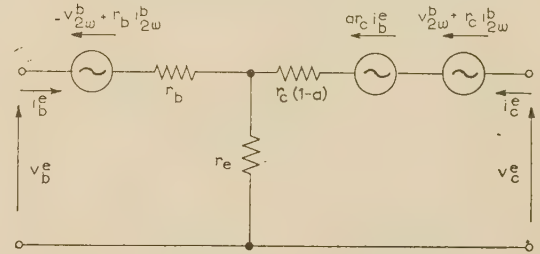


Fig. 12.—T-network equivalent circuit of the common-emitter configuration with second-harmonic voltage generators derived from the common-base case.

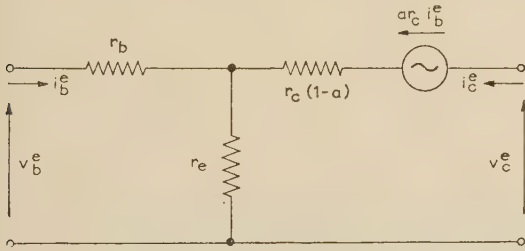


Fig. 10.—Small-signal T-network equivalent circuit of the common-emitter configuration.

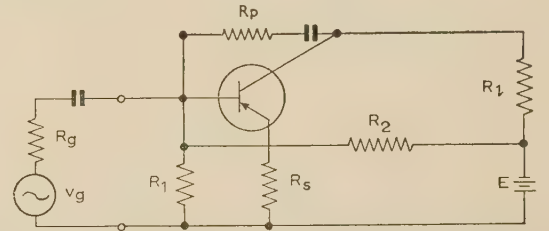


Fig. 13.—Common-emitter amplifier with negative feedback.

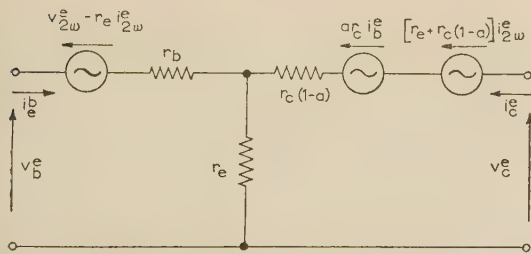


Fig. 11.—T-network equivalent circuit of the common-emitter configuration with second-harmonic voltage generators.

In Fig. 11 the generators including \$v_{2\omega}^b\$ and \$i_{2\omega}^b\$ can be replaced by generators including \$v_{2\omega}^b\$ and \$i_{2\omega}^b\$ by using eqn. (42). After some calculation the equivalent circuit shown in Fig. 12 is then obtained.

(C) SINGLE-STAGE AMPLIFIER WITH NEGATIVE FEEDBACK

As an example of the use of the T-networks developed, the distortion factor of the output circuit for the common-emitter with negative feedback shown in Fig. 13, where \$R_p\$ and \$R_s\$ cause the feedback, will be calculated.

By omitting the bias network (i.e. \$R_1\$ and \$R_2\$) of Fig. 13 and using Fig. 12 as the equivalent of the transistor, the equivalent circuit for the amplifier becomes that shown in Fig. 14.

Using the amplitudes of currents and voltages the solution of

The following inequalities will be fulfilled by practical amplifiers with junction transistors:

$$\begin{aligned} r_e + r_b(1 - a) &\ll \begin{cases} R_g \\ R_l \\ r_c(1 - a) \\ R_p \end{cases}; \\ R_p, R_g, r_b, R_l &\ll r_c; R_s \ll R_p; \frac{a}{1 - a} \gg 1. \end{aligned} \quad (49)$$

By the use of these inequalities, eqn. (48) is reduced to

$$i_{c2}^e \approx \frac{v_{2\omega}^b (R_p + R_g) + R_p r_c i_{2\omega}^b [r_e + R_s + r_b(1 - a) + R_g]}{D'} \quad (50)$$

where

$$D' = [r_e + R_s + r_b(1 - a)](R_p + R_g + R_l) + R_g[R_l + R_p(1 - a)]$$

The terms $v_{2\omega a}^b$ and $i_{2\omega a}^b$ are given by eqns. (22) and (23). By analogy with the principles used in calculating the distortion

From eqns. (55) and (58) the relation between K_{df}^e and K_d^e is found to be

$$F_r = \frac{k_{df}^e}{k_d^e} \simeq \frac{\left(1 + \frac{R_g}{R_p}\right)(K_1^b + K_2^b R_l + K_3^b R_l^2) + (K_4^b + K_5^b R_l + K_6^b R_l^2)(R_g + R_s + h_{11}^b)}{K_1^b + K_2^b R_l + K_3^b R_l^2 + (K_4^b + K_5^b R_l + K_6^b R_l^2)(R_g + h_{11}^b)} \times \frac{R_l \Delta_h^b + h_{11}^b + R_g(1 + h_{21}^b + \frac{h_{22}^b R_l}{R_p})}{h_{11}^b + R_s + R_g(1 + h_{21}^b + \frac{R_l}{R_p})} \quad (59)$$

factor for the common-base amplifier, all currents and voltages in the expressions for $v_{2\omega a}^b$ and $i_{2\omega a}^b$ are expressed by i_{c1}^e . Using conventional linear-network calculations, the following approximate expressions are obtained:

$$i_{c1}^b \simeq -\left(1 + \frac{R_l}{R_p}\right)i_{c1}^e \text{ and } v_{c1}^b \simeq -R_l\left(1 + \frac{R_s}{R_l}\right)i_{c1}^e \quad (51)$$

where the inequalities (49) have been used.

Substitution of eqns. (22), (23) and (51) in eqn. (50) yields.

$$i_{c2}^e \simeq \frac{K_1^b\left(1 + \frac{R_l}{R_p}\right)^2 + K_2^b\left(1 + \frac{R_l}{R_p}\right)\left(1 + \frac{R_s}{R_l}\right)R_l + K_3^b R_l^2\left(1 + \frac{R_s}{R_l}\right)^2}{D'} (R_p + R_g)(i_{c1}^e)^2 + \frac{\left[K_4^b\left(1 + \frac{R_l}{R_p}\right)^2 + K_5^b\left(1 + \frac{R_l}{R_p}\right)\left(1 + \frac{R_s}{R_l}\right)R_l + K_6^b R_l^2\left(1 + \frac{R_s}{R_l}\right)^2\right][r_e + R_s + r_b(1 - a) + R_g]R_p}{D'} (i_{c1}^e)^2 \quad (52)$$

From eqn. (52) the distortion factor of the output current can be found as

$$k_{df}^e = \frac{|i_{c2}^e|}{|i_{c1}^e|} \quad (53)$$

The following inequalities will be satisfied for the majority of practical amplifiers:

$$R_s \ll R_l \ll R_p \quad (54)$$

and by using eqn. (54), $a \simeq -h_{21}^b$, and $r_e + r_b(1 - a) \simeq h_{11}^b$, an approximate form of eqn. (53) can be obtained:

$$k_{df}^e = \frac{|i_{c2}^e|}{|i_{c1}^e|} \simeq \frac{\left(K_1^b + K_2^b R_l + K_3^b R_l^2\right)\left(1 + \frac{R_g}{R_p}\right) + (K_4^b + K_5^b R_l + K_6^b R_l^2)(h_{11}^b + R_s + R_g)}{D''} i_{c1}^e \quad (55)$$

where $D'' = h_{11}^b + R_s + R_g\left(1 + h_{21}^b + \frac{R_l}{R_p}\right)$

It is interesting to compare this expression with that for the distortion factor of the common-emitter amplifier without feedback as given in eqn. (43).

Eqns. (26) and (27) lead to the approximate values

$$\left. \begin{aligned} K_A &\simeq K_1^b + K_2^b R_l + K_3^b R_l^2 \\ K_B &\simeq K_4^b + K_5^b R_l + K_6^b R_l^2 \end{aligned} \right\} \quad (56)$$

it being assumed that

$$R_l \ll \frac{1}{h_{22}^b} \simeq r_c \text{ and } -h_{21}^b \simeq a \simeq 1$$

and eqn. (40) similarly leads to

$$\begin{aligned} h_{21}^e &\simeq \frac{h_{21}^b}{1 + h_{21}^b}; \quad h_{22}^e \simeq \frac{h_{22}^b}{1 + h_{21}^b}; \\ h_{11}^e &\simeq \frac{h_{11}^b}{1 + h_{21}^b} \text{ and } \Delta_h^e \simeq \frac{\Delta_h^b}{1 + h_{21}^b} \end{aligned} \quad (57)$$

Substituting eqns. (56) and (57) in eqn. (43) yields

$$k_d^e \simeq \frac{K_1^b + K_2^b R_l + K_3^b R_l^2 + (K_4^b + K_5^b R_l + K_6^b R_l^2)(R_g + h_{11}^b)}{R_l \Delta_h^b + h_{11}^b + R_g(1 + h_{21}^b + \frac{h_{22}^b R_l}{R_p})} i_{c1}^e \quad (58)$$

Eqn. (59) can be further simplified only if the following inequalities are satisfied:

$$R_s \ll R_l; \quad R_g \ll R_p; \quad h_{22}^b R_l \ll 1 + h_{21}^b \quad (60)$$

in which case F_r is found to be

$$F_r \simeq \frac{1}{1 + \frac{R_g}{R_g + R_l} \frac{R_l}{R_p} A_i^e - \frac{R_l}{R_g + R_l} \frac{R_s}{R_l} A_v^e} \quad (61)$$

where $A_i^e = h_{21}^e$, $A_v^e = -R_l/h_{11}^e$, and $R_l^e = h_{11}^e$ are the approximate expressions for the current amplification, voltage amplification and input resistance, respectively, of the common-emitter amplifier without feedback when expressions (49) and (60) are satisfied.

(6) EXPERIMENTAL VERIFICATION

For the experimental verification of the formulae developed for the distortion factor of the output current, some typical junction transistors with the following characteristic data were used: maximum collector dissipation about 50 mW, and α cut-off frequency about 500 kc/s. The expressions were checked for the common-base configuration (Fig. 5), the common-emitter configuration without external feedback (Fig. 6), and the common-emitter with negative feedback (Fig. 13), using the same transistor in the three cases. The corresponding expressions are (33), (43) and (53), respectively. To compare the measured values of the distortion factor with the calculated values, it is necessary to determine the following constants of the transistor at the working point chosen:

$$r_e, r_b, r_c, a, h_{11}, h_{12}, h_{21}, h_{22}, K_1, K_2, K_3, K_4, K_5, K_6$$

The h parameters were measured by conventional methods as described in the literature,^{2,3} and the T-network parameters were derived from the h parameters. The K parameters can be

determined either directly from the defining equations (22) and (23) by measuring the parameters at conveniently chosen neighbouring points, or indirectly from expression (33) by measuring the distortion factor for six different values of R_g and R_l . Both methods were used, and, allowing for experimental uncertainty, the results were found to agree satisfactorily.

Table 1

MEASURED AND DERIVED DATA

$V_c^b = -2.7$ volt; $I_e^b = 4.0$ mA; $f = 1000$ c/s

h_{11}^b , ohms	13.6
h_{12}^b	1.29×10^{-3}
h_{21}^b	-0.982
h_{22}^b , mhos	2.09×10^{-6}
h_{11}^e , ohms	821
h_{12}^e	4.4×10^{-4}
h_{21}^e	59.4
h_{22}^e , mhos	1.26×10^{-4}
a	0.983
r_e , ohms	3.4
r_b , ohms	620
r_c , kilohms	480
K_1^b ohms/mA	-0.42
K_2^b ohms/volt	0.082
K_3^b volt $^{-1}$	0.052×10^{-3}
K_4^b mA $^{-1}$	3.8×10^{-5}
K_5^b volt $^{-1}$	29×10^{-5}
K_6^b mhos/volt	10.3×10^{-8}

The measured and derived data are given in Table 1, where the K parameters were determined by the direct method using

$$V_c^b \pm 0.5 \text{ volt and } I_e^b \pm 0.5 \text{ mA}$$

as neighbouring points.

Using the indirect method, the distortion factor was measured for the values of R_g and R_l given in Table 2.

Table 2

VALUES OF LOAD- AND GENERATOR-RESISTANCE USED FOR THE DETERMINATION OF THE K PARAMETERS

$R_g, k\Omega$	40	40	40	1×10^{-3}	1×10^{-3}	1×10^{-3}
$R_l, k\Omega$	0.5	5	10	0.1	5	10

From expressions (33) and (43) it can be seen that the determining terms for the distortion factor are K_A and K_B as given in eqns. (26) and (27). Using the approximations

$$h_{21}^b \simeq -1 \text{ and } h_{22}^b R_l \ll 1$$

K_A and K_B can be written

$$K_A \simeq K_1^b + K_2^b R_l + K_3^b R_l^2$$

$$K_B \simeq K_4^b + K_5^b R_l + K_6^b R_l^2$$

It has been the practice, in treating non-linear distortion in junction transistors, to consider only the variation of input resistance with emitter current, i.e. in the above expression for K_A to neglect the terms $K_2^b R_l$ and $K_3^b R_l^2$ relative to K_1^b . This is correct only for rather small values of R_l for the chosen working point, as can be seen from Table 1 and it is furthermore possible

to obtain a compensation between the terms in the expression for K_A , as the two last terms are of opposite sign to the first one.

The variation of K_1^b with emitter current is approximately proportional to I_e^{-2} , and for large emitter currents the two last terms in the expression for K_A become even more important. The first two terms in the above expression for K_B are related to the variation of the short-circuit current gain with emitter current and collector voltage. With $R_l = 1$ kilohm, all three terms in the expression for K_B are usually of the same order of magnitude, except over the range near the maximum of h_{21}^b as a function of I_e^b , where the term K_4^b can be neglected.

The distortion factor of the output current was measured with the transistor in the common-base configuration with a constant load resistance and constant output current for different values of the generator resistance R_g at a frequency of 1000 c/s. The experimental results, shown in Fig. 15, are seen to be in satisfactory agreement with those calculated from expression (33).

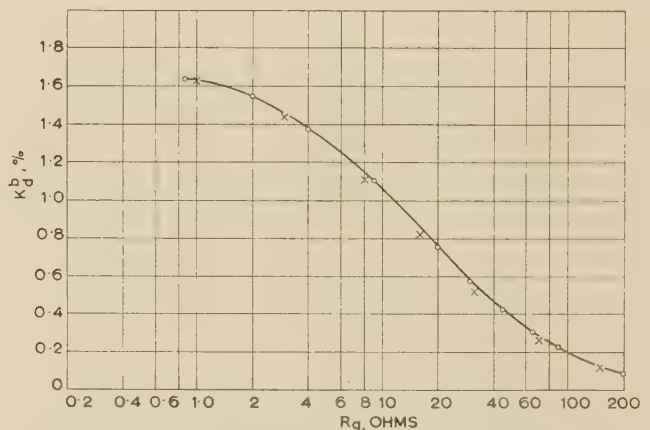


Fig. 15.—Output distortion factor of common-base amplifier as a function of the internal resistance of the voltage source.

○ Experimental points.

× Calculated points.

$R_l = 1$ kilohm; $I_e^b = 0.85$ mA; $f = 1000$ c/s.

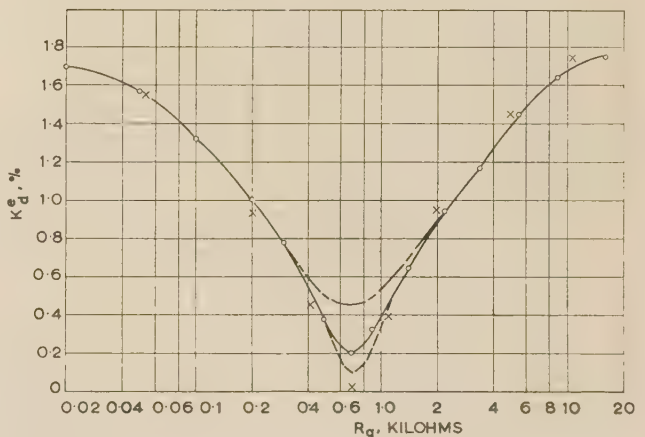


Fig. 16.—Output distortion factor of common-emitter amplifier as a function of the internal resistance of the voltage source.

○ Experimental points } $f = 1000$ c/s.

× Calculated points } experimental.

--- $f = 500$ c/s

--- $f = 2000$ c/s

$R_l = 1$ kilohm; $I_e^b = 0.85$ mA.

The same measurements were made with the transistor in the common-emitter configuration, and the results are shown in Fig. 16, where the theoretical points are calculated from expression (43).

It is to be noted that the distortion factor in Fig. 16 has a well-defined minimum for a special value of R_g , as was to be expected since in expression (43) the two terms in the numerator have different signs, while for $R_g = 0$ the first term is numerically the greater. It is seen that the experimental curve follows the theoretical one except in the neighbourhood of the minimum, which is explained by the fact that the transistor parameters are only approximatively frequency-independent and real-number constants. The frequency dependence is illustrated in Fig. 16 by the two experimental curves measured at frequencies of 500 c/s and 2000 c/s.

The condition of minimum distortion for the simple common-emitter amplifier can be found from eqns. (38) and (43):

$$R_{g(min)} = h_{21}^e \frac{K_C}{K_D} - h_{11}^e \quad . \quad . \quad . \quad (62)$$

or
$$R_{g(min)} = \frac{-h_{21}^e K_A}{(1 + h_{21}^e) K_B} - \frac{h_{11}^e}{1 + h_{21}^e} \quad . \quad . \quad . \quad (63)$$

where K_A , K_B , K_C and K_D are functions of R_1 as given in eqns. (26), (27) and (38). The value of $R_{g(min)}$ in the example shown in Fig. 16 is of the same order of magnitude as the input resistance of the amplifier.

The distortion factor of the output current for the common-emitter amplifier with negative feedback also has a distinct minimum when R_g is varied. The experimental and theoretical

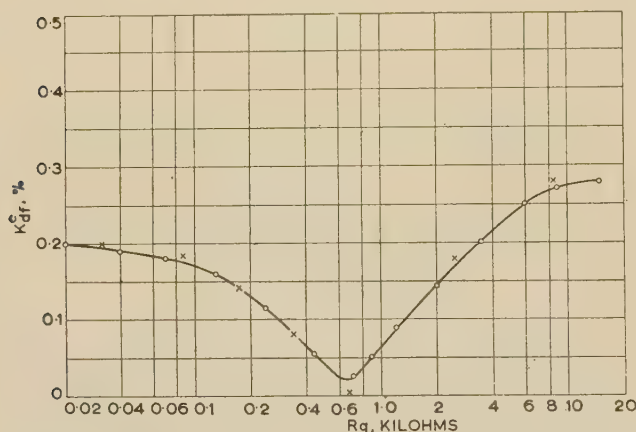


Fig. 17.—Output distortion factor of common-emitter amplifier with negative feedback as a function of the internal resistance of the voltage source.

○ Experimental points.
 × Calculated points.
 $R_1 = 1$ kilohm. $i_{c1}^e = 0.85$ mA.
 $R_p = 10$ kilohms. $f = 1000$ c/s.
 $R_s = 0.1$ kilohm.

points for this case are shown in Fig. 17 for the same output levels as in the two other cases.

In the above-mentioned experiments, the magnitude of the third-harmonic distortion was found to be approximately ten times less than that of the second-harmonic distortion, except in the neighbourhood of the minima where the distortions were of the same order of magnitude. It should be mentioned that the third-harmonic distortion proved to be less dependent on R_g and, unlike that of the second-harmonic, it exhibited no minima for specific values of R_g .

(7) CONCLUSIONS

It has been shown by an analytical investigation of non-linear distortion in junction-transistors that the influence of the variation of all the small-signal parameters may be important in many cases depending on the working conditions of the transistor. Most of the earlier workers in this field have considered the variation of only a few of these parameters, especially the variation of the input resistance with emitter current.

It is thought that the principles developed here may be applied with advantage in more detailed investigations of non-linear distortion in many types of transistor amplifiers.

It has been further shown that it is possible to obtain a considerable reduction of the output distortion in the simple common-emitter amplifier by choosing the working conditions in such a way that the input and output distortions of the transistor cancel each other. This can be brought about only at frequencies so low that the transistor parameters can be considered real and frequency independent, a fact which may seriously restrict the practical utility of the effect. With a view to practical applications, the variation in the distortion constants of transistors of identical types is also an important problem which requires attention.

(8) ACKNOWLEDGMENTS

The author wishes to express his gratitude for the advice and encouragement of Professor J. Oskar Nielsen and for the facilities provided at Teleteknisk Forskningslaboratorium during the progress of the work described in the paper.

(9) REFERENCES

- (1) FELDTKELLER, R., and WOLMAN, W.: 'Fastlineare Netzwerke', *Veröffentlichungen aus dem Gebiete der Nachrichtentechnik*, 1931, 4, p. 281.
- (2) SHEA, R. F.: 'Principles of Transistor Circuits' (J. Wiley and Sons, 1953).
- (3) LO, A. W., ENDRES, R. O., ZAWELS, J., WALDHauer, F. D., and CHENG, C. C.: 'Transistor Electronics' (Prentice-Hall Inc., 1955).

INTERMODULATION IN BRIDGE DETECTOR AMPLIFIERS

By G. J. JOHNSON, B.Sc., and A. M. THOMPSON, B.Sc.

(The paper was received 11th July, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

A discussion is given of the effects of intermodulation in bridge detector amplifiers and the means by which they may be minimized. A figure of merit is defined to assess the performance of an amplifier, and is used to demonstrate the importance of tuning and feedback. The problems associated with the use of amplifiers containing both tuned and untuned stages are discussed, and the arrangement of these stages for minimum intermodulation is derived. The paper includes a description of a practical amplifier for use at 1 and 1.6 kc/s in which the intermodulation is unusually low.

LIST OF PRINCIPAL SYMBOLS

- i_a = Anode current.
 v = Grid-cathode voltage.
 v_o = Output voltage.
 v_i = Input voltage.
 V_f = Equivalent fundamental input voltage at input of a single stage.
 V_{fn} = Equivalent fundamental input voltage at input of n stages.
 R = Resistance.
 C = Capacitance.
 Z = General impedance.
 Z_ω = Impedance at angular frequency ω .
 \bar{Z}_ω = Complex conjugate of Z_ω .
 ω = Angular frequency.
 m_u = Gain of an untuned stage.
 m_h = Gain of a tuned stage at harmonic frequencies.
 m_f = Gain of a tuned stage at the fundamental frequency.
 M = Figure of merit, without feedback.
 M_f = Figure of merit, with feedback.
 a, b, a_1, b_1, a', b' = Constants.
 β = Fraction of output voltage fed back to the input.
 β_ω = Fraction of output voltage of frequency ω , fed back to the input.

the design of which presents several problems. The limiting sensitivity set by thermal noise may be improved by the use of low-noise valves and narrow-bandwidth circuits, but errors caused by non-linearity in the amplifier may be a major limitation on the accuracy attainable. The paper investigates the magnitude of these errors and considers means for reducing them.

(2) NON-LINEARITY IN THE DETECTOR AMPLIFIER

The production of harmonic voltages in the detector amplifier is not troublesome, since such voltages vanish when the bridge is balanced. However, the balance conditions of all bridges are frequency sensitive to some extent, so that, unless the bridge supply is pure and the bridge components are linear, there will be harmonic voltages present at the input to the detector amplifier. Intermodulation in the amplifier will produce fundamental voltage from adjacent harmonics, and the bridge must be set to a false balance to produce a fundamental null at the output.

(3) FIGURE OF MERIT

To compare amplifiers from the aspect of intermodulation it is desirable to be able to associate with a particular amplifier a number which increases as the intermodulation decreases. It will be shown that the voltage at the difference frequency produced by intermodulation is proportional to the product of the amplitudes of the voltages concerned. Hence, a figure of merit is defined as the ratio of the product of the input amplitudes of the two harmonics considered to the fundamental input voltage required to produce the same fundamental output voltage, all quantities being measured in volts.

(3.1) Triode Amplifier

The anode current of a triode is a function of both the grid and the anode voltages. Llewellyn¹ has shown that for an input voltage

$$v = V_2 \cos 2\omega t + V_3 \cos 3\omega t \quad \dots \quad (1)$$

the component of anode current at the difference frequency is

$$i_{a(3\omega - 2\omega)} = \frac{\frac{1}{2} \left[-\mu^2 R R' + \mu \frac{\partial \mu}{\partial V_p} (R^2 - Z_{3\omega} \bar{Z}_{2\omega}) + \frac{\partial \mu}{\partial V_g} (R + Z_{3\omega})(R + \bar{Z}_{2\omega}) \right] V_2 V_3}{(R + Z_{3\omega})(R + \bar{Z}_{2\omega})(R + Z_{3\omega - 2\omega})} \quad (2)$$

(1) INTRODUCTION

For impedance measurements at audio frequencies bridge methods are widely used. With suitable unbalance detectors these methods are capable of high precision, being limited only by the bridge components themselves and the characteristics of the detector used. This detector is usually a valve amplifier,

where μ = Amplification factor.

R = Anode resistance.

$R' = \partial R / \partial V_p$.

$Z_{3\omega}$ = Anode load impedance at frequency 3ω .

$\bar{Z}_{2\omega}$ = Complex conjugate of the impedance at frequency 2ω .

If eqn. (2) is divided by $\mu/(R + Z_{3\omega - 2\omega})$, the equivalent fundamental input voltage required to produce this output is obtained, namely

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Johnson and Mr. Thompson are at the National Standards Laboratory, Commonwealth Scientific and Industrial Research Organization, Australia.

$$V_f = \frac{1}{2} \frac{\left[-\mu^2 R R' + \mu \frac{\partial \mu}{\partial V_p} (R^2 - Z_{3\omega} \bar{Z}_{2\omega}) + \frac{\partial \mu}{\partial V_g} (R + Z_{3\omega})(R + \bar{Z}_{2\omega}) \right] V_2 V_3}{\mu(R + Z_{3\omega})(R + \bar{Z}_{2\omega})} \quad (3)$$

The figure of merit for the triode is then

$$M = \frac{2\mu(R + Z_{3\omega})(R + \bar{Z}_{2\omega})}{\left[-\mu^2 R R' + \mu \frac{\partial \mu}{\partial V_p} (R^2 - Z_{3\omega} \bar{Z}_{2\omega}) + \frac{\partial \mu}{\partial V_g} (R + Z_{3\omega})(R + \bar{Z}_{2\omega}) \right]} \quad (4)$$

It is interesting to note that this figure of merit for the triode depends only on the anode load impedances at the harmonic frequencies and not on that at the fundamental frequency.

The performance of a triode is completely specified by eqn. (4), but it is not easy to appreciate the effects of the various parameters unless some numerical values are considered. It is convenient first to study the case when the load is a pure resistor. For a triode-connected EF 86 with an anode voltage of 200 volts and grid voltage of -3 volts the following values were measured:

$$\frac{\partial \mu}{\partial V_g} = 1.3 \text{ per volt}, \quad \mu = 37, \quad R = 14 \text{ kilohms},$$

$$R' = -98 \text{ ohms/volt}, \quad \frac{\partial \mu}{\partial V_p} = 0.02 \text{ per volt}.$$

Substitution of these values into eqn. (4) gives the following values of M :

Z/R	0	1	2	4	∞
M	6.3	20	35	60	130

It can thus be seen that the figure of merit for a triode increases as the load resistance increases.

(3.2) Pentode Amplifier

The ideal pentode differs from the triode in that the anode current depends only on the grid voltage. This implies that the anode resistance is infinite, which for finite transconductance requires an infinite amplification factor. Hence the figure of merit for the pentode may be derived by letting $R \rightarrow \infty$, $\mu \rightarrow \infty$, $\mu/R \rightarrow g_m$ where g_m is the transconductance of the valve. To do this it is convenient to transform eqn. (4) to

$$M = \frac{2(1 + Z_{3\omega}/R)(1 + \bar{Z}_{2\omega}/R)}{-g_m R' + \frac{\partial \mu}{\partial V_p} (1 - Z_{3\omega} \bar{Z}_{2\omega}/R^2) + \frac{1}{\mu} \frac{\partial \mu}{\partial V_g} (1 + Z_{3\omega}/R)(1 + \bar{Z}_{2\omega}/R)} \quad (5)$$

which for $R \rightarrow \infty$, $\mu \rightarrow \infty$, becomes

$$M = \frac{2}{-g_m R' + \frac{\partial \mu}{\partial V_p}} \quad (6)$$

and putting $\mu = g_m R$, and $\frac{\partial g_m}{\partial V_p} = -\frac{1}{\mu} \frac{\partial g_m}{\partial V_g}$ gives

$$M = \frac{2g_m}{\partial g_m / \partial V_g} \quad (7)$$

It is instructive to derive eqn. (7) by expanding the anode current as a series of powers of the grid voltage. Then retaining only second-order terms,

$$i_a = av + bv^2 \quad (8)$$

where

$$b = \frac{1}{2} \partial g_m / \partial V_g$$

Substituting $v = V_2 \cos 2\omega t + V_3 \cos 3\omega t$ into eqn. (8) and referring to the input by dividing by the gain gives

$$V_f = \frac{b V_2 V_3}{a}$$

where V_f is the equivalent input which would produce the actual output at frequency ω .

Hence

$$M = a/b = \frac{2g_m}{\partial g_m / \partial V_g} \quad (7)$$

For the valve used in the previous measurements (EF86), connected as a pentode with the anode current corresponding to zero anode load, the figure of merit for the pentode was 7.2. That this is of the same order as that for a triode with zero anode load was perhaps to be expected, since under these conditions the triode behaves essentially as a pentode in that its anode current is determined only by the grid voltage.

(4) EFFECTS OF TUNING

(4.1) Triode

When the load in the anode circuit of a triode is complex, a general discussion of eqn. (4) is difficult. It is possible, however, to consider a particular case of some practical importance. Suppose that the anode load is a parallel circuit tuned to the frequency ω_0 . Then, assuming that all the losses may be represented by a resistance R_s in shunt with the circuit, the impedance can be written

$$Z = \frac{1}{\omega_0 C j \left(\frac{\omega}{\omega_0} \right)} \frac{1}{1 - \left(\frac{\omega}{\omega_0} \right)^2 + \frac{j}{Q} \left(\frac{\omega}{\omega_0} \right)} \quad (9)$$

where $Q = R_s \omega_0 C$. For reasonably large values of Q this may be more simply expressed as

$$Z = \frac{1}{\omega_0 C j \left(\frac{\omega}{\omega_0} \right)} \frac{1}{1 - \left(\frac{\omega}{\omega_0} \right)^2} \quad (10)$$

This value for Z has been substituted in eqn. (4) with $\omega_0 = 10^4$ rad/sec and the figures given above for the EF 86, the corresponding value of M being plotted in Fig. 1. The figure of merit increases rapidly with decrease of tuning capacitance up to a maximum of 130 for zero capacitance; however, it is within 10% of its maximum value for a capacitance of $200 \mu\mu\text{F}$. Inductances with distributed capacitance less than this and with sufficiently high Q -factors at audio frequencies are conveniently made using quite small coils with Ferroxcube cores.

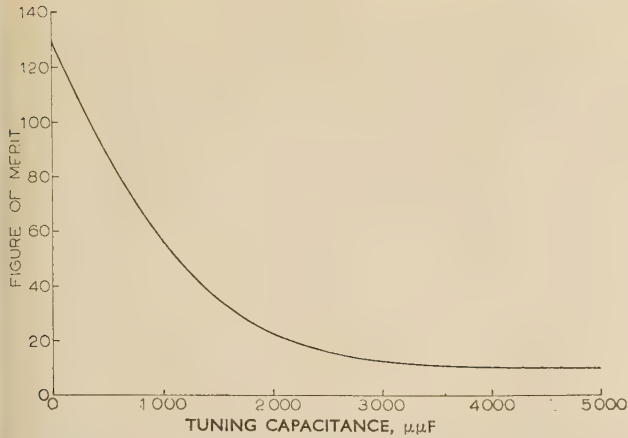


Fig. 1.—Variation of figure of merit with tuning capacitance in the anode circuit of a triode.

(4.2) Pentode

Since it has been assumed that the anode current of a pentode is independent of the anode voltage, tuning of the plate circuit can have no effect on the intermodulation current which flows. Consequently the equivalent fundamental input voltage is independent of the anode load impedance.

Although the anode load of a pentode does not directly affect the intermodulation distortion of a stage, there are usually indirect effects which make the use of tuned anode loads desirable. Tuned circuits provide low gain at the harmonic frequencies while maintaining high gain at the fundamental, thus reducing the level of the harmonics at the succeeding stages. Furthermore, from a given d.c. power supply they enable higher transconductance and consequently higher figures of merit to be achieved.

(5) EFFECTS OF FEEDBACK

For simplicity, the succeeding discussion will use the pentode coefficients, although a similar argument may be applied to the triode case if the more complicated coefficients given in Section 3.1 are used.

(5.1) Untuned Amplifier

For an amplifier whose anode load is a pure resistance R_L , the output voltage may be written

$$v_o = R_L(av + bv^2) \quad (11)$$

where terms higher than the second degree have been neglected. If the input voltage v_i has added to it a voltage derived from the output, then

$$v = v_i + \beta v_o \quad (12)$$

where β is a fraction of the output voltage added to the input. Substitution for v_o from eqn. (11) into eqn. (12) gives

$$v = v_i + \beta R_L(av + bv^2) \quad (13)$$

If the output current may be expressed as a power series in the input voltage, it is possible to write

$$av + bv^2 = a_1 v_i + b_1 v_i^2 \quad (14)$$

Substituting for v_i from eqn. (13) and equating coefficients of like powers of v gives

$$a_1 = \frac{a}{1 - a\beta R_L} \quad (15)$$

$$b_1 = \frac{b}{(1 - a\beta R_L)^3} \quad (16)$$

The corresponding figure of merit is then

$$M = \frac{a}{b}(1 - a\beta R_L)^2 \quad (17)$$

Eqn. (17) indicates the effectiveness of feedback, for the figure of merit with feedback is given by the product of the figure of merit without feedback and the square of the gain reduction factor. It is possible to rewrite eqn. (17) as

$$M = \frac{a}{b} \left(\frac{\text{gain without feedback}}{\text{gain with feedback}} \right)^2 \quad (18)$$

where the importance of a high gain without feedback is immediately apparent. For a certain required gain the figure of merit is proportional to the product of a/b and the square of the gain without feedback.

Measurements on an EF80 valve with 1 kc/s as the fundamental frequency and equal 2 and 3 kc/s voltages showed a linear relationship between M and the square of the ratio of the gain without feedback to the gain with feedback, the slope of the graph being 2.72. The measured transconductance was 3 mA/volt, while the coefficient b was 1.1 mA/volt², giving $a/b = 2.73$.

(5.2) Frequency-Selective Feedback

Suppose that, as before, the input voltage is that given by eqn. (1) and that in the anode circuit there is an impedance whose value is $Z_{2\omega}$ at frequency 2ω and $Z_{3\omega}$ at frequency 3ω . Furthermore, let the feedback factors at the two frequencies be $\beta_{2\omega}$, $\beta_{3\omega}$, respectively; then the grid-cathode voltages are

$$v_{2\omega} = \frac{V_2}{1 - a\beta_{2\omega}Z_{2\omega}} \quad (19)$$

at frequency 2ω

$$\text{and} \quad v_{3\omega} = \frac{V_3}{1 - a\beta_{3\omega}Z_{3\omega}} \quad (20)$$

at frequency 3ω .

Hence the figure of merit is

$$M_f = \frac{a}{b}(1 - a\beta_{2\omega}Z_{2\omega})(1 - a\beta_{3\omega}Z_{3\omega})$$

$$M_f = \frac{a}{b} \left(\frac{\text{Product of harmonic gains without feedback}}{\text{Product of harmonic gains with feedback}} \right) \quad (21)$$

Eqn. (21) indicates the possibility of achieving a high figure of merit by the use of feedback of the harmonics while leaving the fundamental gain essentially unaffected.

(6) MULTI-STAGE AMPLIFIERS

Most bridge detectors require more than one stage of amplification, and these additional stages will produce intermodulation distortion of their own. It is possible to derive a figure of merit for a general n -stage amplifier all of whose stages are different, but the resultant expressions are not really useful. However, the derivation for two particular amplifiers will suffice to indicate the method of approach and to specify the requirements for these cases.

(6.1) n Identical Stages

Consider first an n -stage amplifier all of whose stages are identical, and suppose that each stage has a gain of m_f at the fundamental and m_h to the harmonics. Then for an input given by eqn. (1) the first stage will produce an equivalent fundamental input of bV_2V_3/a . At the input to the second stage there will

be $m_h(V_2 \cos 2\omega t + V_3 \cos 3\omega t)$ and consequently an equivalent fundamental at the input to the first stage of

$$\frac{bm_h^2 V_2 V_3}{am_f}$$

Hence from two stages the equivalent fundamental input voltage is

$$V_{f2} = \frac{bV_2 V_3}{a} \left(1 + \frac{m_h^2}{m_f^2}\right) \quad . \quad . \quad . \quad (22)$$

For n stages,

$$V_{fn} = \frac{bV_2 V_3}{a} \sum_{r=0}^{n-1} \frac{m_h^{2r}}{m_f^2} \quad . \quad . \quad . \quad (23)$$

Eqn. (23) indicates the desirability of attenuating the harmonics in the coupling network between stages. Under the best conditions, when m_h^2/m_f^2 is much less than unity, the overall performance of the amplifier is determined by the first stage alone.

(6.2) Two Multi-Stage Amplifiers in Cascade

It is sometimes required to use a certain amount of untuned gain in conjunction with a tuned amplifier, and the question arises whether the tuned stages should precede the untuned ones or vice versa.

Suppose that there are p stages of tuned gain and q stages of untuned gain and that the constants a and b are the same for all stages. Let the gain per stage at the fundamental be m_f in the tuned stages and m_u in the untuned ones, while the gain at the harmonics in the tuned stages is m_h per stage. Then when the tuned stages precede the untuned ones the contribution of fundamental at the input for an applied voltage ($V_2 \cos 2\omega t + V_3 \cos 3\omega t$) is

$$V_{f1} = \frac{bV_2 V_3}{a} \sum_{r=0}^{p-1} \frac{m_h^{2r}}{m_f^2} \quad . \quad . \quad . \quad (24)$$

At the input to the untuned stages the harmonic voltages will be $m_h^p(V_2 \cos 2\omega t + V_3 \cos 3\omega t)$, so that the intermodulation product referred to the input of the tuned stages is

$$V_{f2} = \frac{bV_2 V_3}{a} \frac{m_h^{2p}}{m_f^p} \sum_{s=0}^{q-1} \frac{m_u^{2s}}{m_u^2} \quad . \quad . \quad . \quad (25)$$

Hence the total fundamental voltage referred to the input is

$$V_{fa} = \frac{bV_2 V_3}{a} \left(\sum_{r=0}^{p-1} \frac{m_h^{2r}}{m_f^2} + \frac{m_h^{2p}}{m_f^p} \sum_{s=0}^{q-1} m_u^{2s} \right) \quad . \quad . \quad (26)$$

Similarly, when the untuned stages precede the tuned ones,

$$V_{fb} = \frac{bV_2 V_3}{a} \left(\sum_{s=0}^{q-1} m_u^{2s} + m_u^{2q} \sum_{r=0}^{p-1} \frac{m_h^{2r}}{m_f^2} \right) \quad . \quad . \quad (27)$$

Eqs. (26) and (27) show that the minimum possible equivalent fundamental input voltage is given when all stages are tuned and m_h^2/m_f^2 is much less than unity. This minimum voltage is $bV_2 V_3/a$, contributed by the first stage. When the untuned stages precede the tuned ones this minimum value is multiplied by m_u^{2q} , which represents an enormous deterioration in performance for even a single stage of untuned gain ahead of the tuned gain.

Hence it may be concluded that at all times the tuned stages should precede the untuned ones; that the coupling between successive stages should be arranged if possible to provide attenuation of the harmonics; and that sufficient tuned stages should be used to make the contribution of the second term in eqn. (26) negligible.

(7) DISCUSSION

The previous Sections have indicated several points which require some comment. First, under all circumstances the equivalent fundamental input voltage produced by intermodulation is proportional to the product of the amplitudes of the two harmonics involved. Hence it is desirable to supply the bridge from as pure a voltage source as possible, or at least to eliminate alternate harmonics. This suggests push-pull amplifiers to drive the bridge and the use of filters ahead of the bridge itself. The limit of improvement in this direction is reached when any additional filter introduces as much distortion as is present at its input. At this stage it is profitable to filter the bridge output. As pointed out in Section 6, it is desirable to place the tuning in an amplifier as early as possible; the ultimate in this direction being tuning in the input circuit when this is possible.

Secondly, the figure of merit for pentodes increases as the ratio a/b increases. Since b is essentially constant over the usual operating range, this implies use of the highest transconductance possible for a given valve. This of itself does not imply that high-transconductance valves are desirable, since it is the ratio a/b that is important. For a pentode-connected 6BR7 the figure of merit is 6.7 at a transconductance of 1.5 mA/volt, whilst for an EF80 the corresponding value is 5.0 for a transconductance of 7 mA/volt. For most bridge detector amplifiers the input valve must be a low-microphony type; these usually have low transconductance, so it is gratifying that their figures of merit are not necessarily inferior to those of the high-transconductance types. However, if feedback is applied in accordance with the conclusions reached in Section 5, the higher-transconductance valves will enable a higher loop gain to be achieved and consequently a higher figure of merit than for the lower-transconductance types. The triode case is not so simple, although one general point which emerges from Section 3 is the desirability of high anode loads for low distortion. For a triode the amplification factor is reasonably constant, and eqn. 4 indicates that, for a given μ and anode load, the figure of merit improves as the anode resistance decreases. This implies the use of a high transconductance, as for the pentode.

Thirdly, a comparison between triodes and pentodes must take into account the effect of feedback as given in eqns. (18) and (21). The maximum value of M without feedback for the triode quoted in Section 3.1 is 130, while for the pentode the corresponding value is about 7. However, it is possible to achieve much higher values of gain with pentodes. A 20-fold increase in gain would, with feedback, give a 400-fold improvement in the figure of merit, so that it is generally possible to design a pentode stage to have a performance superior to that of a triode stage.

(8) A PRACTICAL AMPLIFIER

The circuit of an amplifier designed for use at 1000 and 1592 c/s is given in Fig. 2. The total current drawn from the 67½-volt h.t. battery is 3 mA, while the heaters are supplied from a 6-volt accumulator. The overall gain of the amplifier is approximately 39 000 at 1000 c/s, while 6 mV of 2 and 3 kc/s voltages together produce the equivalent of 1 μ V of 1 kc/s at the input.

The first two stages operate as cathode-followers with series tuned circuits across the cathode resistors. This arrangement has several practical advantages. Each stage operates with 100% negative feedback and hence produces only small amounts of intermodulation distortion. The useful gain at the fundamental of approximately Q is produced by the series tuned circuit, whilst the gains at 2 and 3 kc/s are only $\frac{1}{2}$ and $\frac{1}{3}$ respectively. Furthermore, this form of coupling has the desirable feature that no direct current flows in the inductor. The final

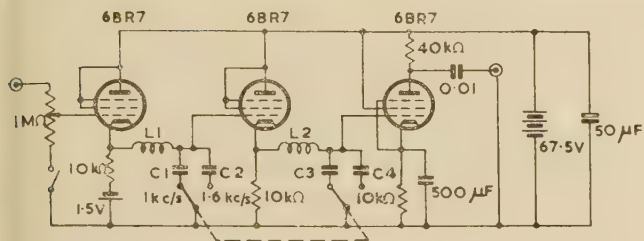


Fig. 2.—Circuit of an amplifier with low intermodulation distortion.

$L1, L2 = 63 \text{ H.}$
 $C1, C3 \approx 400 \mu\text{F.}$
 $C2, C4 \approx 160 \mu\text{F.}$

stage is a conventional flat-tuned stage provided to enable the output to be presented on a general-purpose oscillograph.

A modified version of this amplifier has been made using a 550-volt rectified mains supply with higher operating currents. This amplifier required 38 mV of both 2 and 3 kc/s voltages to produce the equivalent of $1 \mu\text{V}$ at 1 kc/s at the input. The corresponding value quoted by Rayner² is 0.26 mV and that quoted by Clothier and Smith³ is 0.5 mV.

(9) CONCLUSIONS

It has been shown that the intermodulation distortion performance of a bridge detector amplifier may, in the best circumstances, be determined by the first stage. This is only possible if the succeeding stages satisfy a particular condition, which is derived. The performance of a single stage improves as the transconductance and the harmonic feedback increase, and—for triodes—as the anode load increases.

Where it is necessary to use untuned stages in a bridge detector amplifier, it is much superior to precede the untuned stages with the tuned stages.

A complete amplifier, whose intermodulation distortion is considerably lower than that of amplifiers previously described, has been built.

(10) REFERENCES

- (1) LLEWELLYN, F. B.: 'Operation of Thermionic Vacuum Tube Circuits', *Bell System Technical Journal*, 1926, **5**, p. 433.
- (2) RAYNER, G. H.: 'A Selective Detector Amplifier for 10–10 000 c/s', *Journal of Scientific Instruments*, 1953, **30**, p. 17.
- (3) CLOTHIER, W. K., and SMITH, W. E.: 'A Tuned Detector Amplifier for Power Frequency Measurements', *ibid.*, 1955, **32**, p. 67.

ON THE AMPLIFICATION FACTOR OF THE TRIODE

By Professor E. B. MOULLIN, M.A., Sc.D., Past President.

(The paper was first received 10th August, 1955, and in revised form 11th May, 1956. It was published as an INSTITUTION MONOGRAPH in November, 1956.)

SUMMARY

The paper discusses why a triode has the property that any assigned anode current is a function of the anode potential plus a constant times the grid potential. Unlike the classic methods of approach, the paper sets out to find the values which these two voltages must have if and when an assigned anode current is flowing. It starts by reminding the reader that, if a given current is crossing the barrier of a diode, the electrons in that stream propel themselves by means of their own mutual repulsions and are not assisted by the attraction of the positive charge on the anode. The means by which the anode potential operates is through the positive charge which it causes to exist on the cathode, thereby regulating the proportion of the total cathode emission which crosses the barrier. Once this current is assigned, the electrons in it continue to gather kinetic energy until they are brought suddenly to rest by colliding with an anode interposed in their path. The potential of this anode is determined from the kinetic energy they have attained.

If a negatively charged grid is interposed in the path of a given stream of current leaving a barrier, it is shown that this negative charge produces scarcely any electric force in the region between the barrier and the plane of the grid. Accordingly, this negatively charged grid will have scarcely any effect on the motion of the electrons approaching it: its presence will not, *per se*, tend appreciably to change the current which is crossing that barrier. However, once the electrons have passed through the meshes of the negative grid they will experience a strong repulsive force from it and will thereby be accelerated in their path. At any point in that path they will have attained additional kinetic energy from the work put into them by the negative charge on the grid: accordingly, the velocity with which they strike the anode is known and thus the potential which that anode must have is known. By means of the particle dynamics of their flight it is found that, for any assigned current, there is a linear relation between the anode voltage and the negative charge on the grid, placed there by means of a battery connected between the grid and the cathode.

Although the charge on the grid produces a negligible electric force at points in the neighbourhood of the barrier, the force at the surface of the grid wires is large because the grid wires are very thin. Thus the work done from the barrier to the surface of a grid wire is finite, occurring mostly in the near vicinity of the grid wire: it is calculable, by familiar methods, per unit of charge on the grid. The net work from the barrier to a grid wire is the difference between the work against the grid charge and the work against the force caused by the electrons composing the anode current. This difference is the potential of the grid, i.e. the voltage of the grid battery required to maintain an assigned charge on the grid when there is an assigned anode current. This relation between grid charge and grid voltage can then be combined with the previously found linear relation between grid charge and anode voltage, for an assigned current, and from this emerges an expression for the amplification factor of the triode when it is passing an assigned anode current. It turns out to be independent of anode current and to have the value deduced long ago for a triode in which the anode current was vanishingly small, and previously thought to be valid only in this limiting condition. A proof that the amplification factor is independent of current—a fact well known by experiment—is believed to be novel.

(1) INTRODUCTION

The very essence of the electrical behaviour of a thermionic triode is described in the statement that the anode current is a function of the anode voltage plus a constant times the grid voltage, i.e.

$$I_a = f(V_a + \mu V_g) \dots \dots \dots (1)$$

This amazing property was noticed by Van der Bijl^{1,2} and described and reported by him in 1913. The year in which the third electrode was introduced is commonly said to be 1907: thus, recognition of the basic electrical property goes back nearly to the time of the emergence of the grid itself into a high-vacuum thermionic discharge tube. It may well be thought that nothing novel remains to be said about this principle, a full 40 years after its enunciation. Up to a point this must indeed be true; but the present paper attempts to show that there is yet room for some new thought and that the old principle can be set out in a way which has some claim to be more complete, more satisfying, than the old.

In 1937 the present author deduced the long-famous three-halves-power law of the diode³ by the process of following a typical electron from the barrier to the anode, in contrast to the then classical method of deducing the force/distance distribution from a solution of the $\nabla^2(\psi) = -4\pi$ equation: possibly others had done so before him. It corresponds to changing from the Eulerian to the Lagrangian method of solving problems of fluid flow.

This change to the Lagrangian method was a conversion from one correct formulation to another correct formulation of the diode problem: it was a convenience, but no more than that. But when seeking recently to treat the triode in the corresponding manner—and, in fact, succeeding in doing so—it emerged that the classic treatment of the triode is fallacious in principle and unreasonably restricted in result. The paper treats the triode by the Lagrangian method, and in so doing discloses the fallacy, or near fallacy, of the conventional method.

To help the reader to understand the Lagrangian treatment of the triode it may be helpful to recapitulate the method as applied to the diode, as follows. In their transit from cathode to anode the electrons must necessarily cross a closed surface, called the barrier surface, at every point of which the electric force is zero. An electric force at any point within the diode arises from electric charges on the electrodes and from electrons in transit. Since the force at the barrier surface is zero by definition, consideration will show that there must be a positive charge on the cathode equal in amount to the total charge on the electrons which are in motion between the cathode surface and the barrier surface. Further consideration (which will be detailed later) shows that the electric force acting on any typical electron which is en route for the anode arises only from those other electrons which are also en route but are behind the particular electron which has been selected for observation.

Accordingly, consideration by the Lagrangian method discloses the problem as being that of a jet of electrons which is self-

Correspondence on Monographs is invited for consideration with a view to publication.

Prof. Moullin is Professor of Electrical Engineering, University of Cambridge.

propelling. An analogy may not be out of place. Liken any typical electron to a thief pursued by a rabble: he receives his stimulus to run only from the rabble behind him, and is quite unconscious of crowds ahead in the street and of people within the house from which he bolted (they are behind the barrier); he is conscious only of the rate at which fresh pursuers do, in fact, emerge through the barrier. Ultimately the unfortunate thief rushes headlong into what turns out to be a police station and he is put into the cells at once.

The kernel of the whole situation is that it is the existence of the stream itself which causes that stream to flow—a self-propelling system. The circumstances are akin to the movement of a jet of gas, but molecular impacts are replaced by the electric repulsion provided by the electrons between the barrier and any typical electron in transit.

When considering the flow of current from the barrier to the anode in a triode it is natural to wish to maintain, so far as possible, the picture which has been developed for a diode. In the triode the electron stream passes through a grid: actual collision between the travelling electrons and the wires of the grid can be prevented by placing a negative charge on the grid, which is the condition of normal use in a triode. Consideration shows that electric charges on the wires of a cylindrical grid produce very little electric force inside this cylinder and that this radial force, small though it is, must alternate in direction from point to point along the axis of the cylinder. Thus a given stream from the barrier can scarcely be affected by the existence of the grid in that portion of space which lies between the barrier and the grid surface. But in the region of space between the grid surface and the anode the charge on the grid will produce a radial force very nearly the same as if the total charge on the grid had been spread uniformly over the cylindrical grid-surface. It is only after a given electron has shot through a mesh of the grid that it experiences any appreciable indication that the grid exists: then it recognizes its presence as a considerable concentrated charge behind it, a charge, moreover, which does not belong to electrons in radial motion behind it. This negative charge on the grid will accelerate every electron after it has passed through the grid and thereby cause its velocity at any given point of its path to be greater than would have obtained if the charged grid were not there. In particular, the electron will collide with the anode with an enhanced kinetic energy, so that the anode must be connected to a battery of higher voltage than would have obtained for a diode passing the same current. Of course, this battery also determines the net positive charge on the cathode and thereby regulates the position of the barrier. It is by this means, and by this means only, that the anode potential determines and regulates the current reaching it. In the approach which is being used here the current is regarded as the independent variable; then is deduced the anode potential which is appropriate to this particular current.

To start this problem with an assigned current is obviously particularly apposite for studying the mechanism which lies behind the experimental relation that any assigned anode current is a function of $(V_a + \mu V_g)$.

The total electric charge on the wires of the grid will be maintained in position on it by means of an appropriate electric battery (or its equivalent) connected between the grid and the anode; in more common parlance, by applying a suitable potential to the grid.

2.2 PRELIMINARY CONSIDERATIONS: RECAPITULATION OF THE DIODE PROBLEM

Consider a cylindrical diode with anode of radius a (see Fig. 1) and cathode of radius k . At some definite, and calculable,

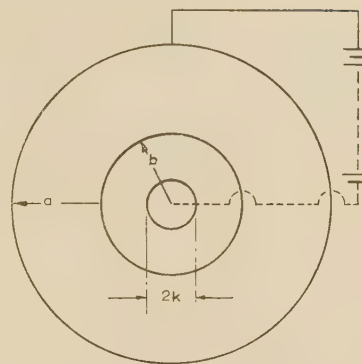


Fig. 1

radius, b , the electric force passes through zero. Inside the cylinder of radius b electrons are flowing both outward and inward: outside this cylindrical surface all electrons are moving radially outward. This cylindrical surface, at which the electric force passes through zero, will be termed the barrier surface. There is little advantage in calling it a potential barrier, for to do so can obscure slightly its primary meaning, which is a surface at which the force passes through zero.

Since the force at radius b is zero and the annulus between b and k contains electrons, it follows that the charge on the surface of k must be positive—notwithstanding that it is connected to the negative end of the battery—and equal in amount to the total charge of all the electrons between b and k .

Since the anode is a metal cylinder, it follows that the electric force must be zero throughout its thickness. This means that there must be a positive charge on the inner surface of the anode, of amount equal to the total charge carried by the electrons in transit between the anode and the barrier. But, because of the inverse-square law, this positive charge on the inside surface of the anode does not produce any electric force at any point inside the anode.

The field of the anode does not penetrate the barrier because it has no field to do so. Similarly, if a grid existed at some radius greater than b , the field of the anode could not penetrate it merely because no anode field existed to do the penetration. Here then is an inexorable fact which shows that there must be some small element of confusion in the time-honoured concept of *Durchgriff*. Thus Möller⁴ defines the factor $1/\mu$ as *Durchgriff*: 'The anode sends through the grid—turns lines of force into the cathode space, which get hold of space-charge electrons and pull them towards the anode.'

Surely this description is equivalent to thinking of a fairy on the anode putting her hand through the grid and getting hold of the electrons, notwithstanding that no lines of force arise from the charge on the anode. And surely it is reasonable to question whether the term *Durchgriff* is really a useful one.

But what has been said of the positive charge on the anode is equally true of the negative charge between any radius r and the anode, in respect of any point inside that radius. Because the electrons are moving radially and the flow has complete symmetry about the axis of the cylinder, the electrons which comprise it can be decomposed into a set of cylindrical shells of uniformly distributed charge. None of these shells produces any electric force at any point inside it. Accordingly, the electric force acting on any typical electron at radius r is provided entirely by the total charge on all the electrons between r and the barrier radius b . It is the space charge behind any typical electron, and that alone, which causes the electron to travel to the anode.

If the diode problem be simplified—in a sense, idealized—to

the extent of postulating that all electrons cross the barrier with the same velocity, then every electron will take the same time, t , to arrive at a distance r from the centre of the cathode. In these circumstances, a quantity of electricity, It , will have crossed unit axial length of the barrier surface after the departure of any typical electron which is in the surface of radius r . Accordingly, the electric force E at radius r is given by the equation

$$E = \frac{2It}{r} \quad (2)$$

Thus the equation of motion is

$$\frac{d^2r}{dt^2} = \frac{2e}{m} \frac{It}{r} \equiv \frac{A'It}{r} \quad (3)$$

This equation has often been solved and leads to the solution

$$I = \frac{2}{9} \left(\frac{2e}{m} \right)^{1/2} \frac{V^{3/2}}{a\beta^2} \quad (4)$$

where a is the radius of the anode and β^2 is a complicated function of the ratio (a/b) , which was evaluated by Langmuir and Blodgett.⁵ The process of the solution is cumbersome and involved; this is almost inevitable, because the symmetry of the system calls for the axis as the origin whereas the equation of motion of an electron calls for an origin on the surface of the barrier.

For this purpose it is helpful to write eqn. (4) in terms of the current density, J , at the surface of the anode, thus:

$$J = \frac{I}{2\pi a} = \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} \frac{V^{3/2}}{a^2\beta^2} \quad (5)$$

$$= \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} \frac{(1 - b/a)^2}{\beta^2} \frac{V^{3/2}}{l^2}$$

where $l \equiv (a - b)$

Let the parameter $\frac{(1 - b/a)^2}{\beta^2}$ be denoted by D^2 , some numerical values of which are collected in Table 1.

Table 1

a/b	1	1.1	2	4	10	20	50	100
D^2	1	0.95	0.89	0.83	0.83	0.84	0.88	0.91

This shows that, when b/a tends to unity, D^2 also tends asymptotically to unity. But even though b/a tends to unity in this formula, the electrodes still remain cylinders and the positive charge on the inside of the outer cylinder still produces no force inside it. Although the radius of curvature of the electrodes tends to infinity, and any unit of area becomes more and more nearly planar, the system never degenerates into a pair of infinite parallel planes. The solution for a pair of infinite parallel planes is very well known, but it seems worth while to recapitulate it here.

Let the traces of the anode and cathode planes be represented by A and C in Fig. 2 and let B be the trace of the barrier plane. Because B is a barrier plane, the electric force at a typical plane P due to the positive charge on the cathode plane C is neutralized by the force at P due to the electron gas between B and C. But the force at P due to the positive charge on the anode is not zero, because the anode is an infinite plane. Herein there is—or appears to be at first sight—a very real difference between the planar and the cylindrical problem, but the effect of this difference vanishes when it is remembered that the positive charge on the

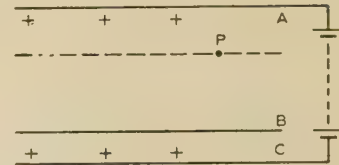


Fig. 2

anode must be equal to the total charge on the electrons in transit between B and A. The force due to a planar distribution of surface density σ is $2\pi\sigma$ at all distances from that plane. Consequently, the upward force at the typical plane P is $2\pi(\sigma - q_1) + 2\pi q_2$, where σ is the charge density on the anode, q_1 is the charge in transit (per unit area of the electrodes) between A and P, and q_2 is the charge in transit between B and P. But since $\sigma = q_1 + q_2$, the total upward force at P is $4\pi q_2$ and thus depends only on the number of electrons in transit behind the typical point P. Of course, this result is obtainable directly from Gauss's theorem, but if this short cut is taken it is not quite so clear how the result comes about.

It has now emerged that, even with electrodes which are planes (and not really pieces of very large cylinders), it is still true to say that the force urging an electron towards the anode can be said to arise from the charge on the electrons which are behind the typical electron whose motion is being considered. Accordingly, the equation of motion is

$$\frac{d^2x}{dt^2} = \frac{4\pi e}{m} Jt \equiv AJt \quad (6)$$

where J is the anode current density and t is the time for electrons to travel a distance x from the barrier, all having crossed it with velocity u_0 . Whence

$$\frac{dx}{dt} = u_0 + \frac{AJt^2}{2} \quad (7)$$

and

$$x = u_0 t + \frac{AJt^3}{6} \quad (8)$$

$$= \frac{t}{3}(u + 2u_0) \quad (9)$$

If $u_0 = 0$, then $t = \left(\frac{6x}{AJ} \right)^{1/3}$ and $u = \frac{(AJ)^{1/3}}{2} (6x)^{2/3}$. On writing $\frac{1}{2}mu^2 = eV$, the following relationships are obtained:

$$V = 3\pi(Jt)x \quad (10a)$$

$$V = \frac{m}{8e} (AJ)^{2/3} (6x)^{4/3} \quad (10b)$$

$$J = \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} \frac{V^{3/2}}{x^2} \quad (10c)$$

Eqn. (10c) is the same as the limiting form of eqn. (5) when b/a tends to unity. Accordingly, planar electrodes lead to the same result as the asymptotic limit for cylindrical electrodes. This result inspires confidence in the validity of treatment by planar electrodes—a treatment which is virtually obligatory as a means of avoiding hopelessly toilsome algebra. The result is well known and the raising of it here may well appear to be mere pedantry; but in a solution which depends almost entirely on thinking clearly about the circumstances obtaining in a certain real problem which is being described analytically, it is wise to ascertain that the mathematical simulation does not contain an inherent flaw.*

* For a more detailed development of eqns. (6) and (10) see Reference 6.

Although the previous discussion has been developed for electrodes which are concentric cylinders, consideration suggests that the general argument is not thereby limited in principle. Obviously it remains unchanged for a spherical diode and it has been shown that it is unchanged for a planar diode (i.e. one in which the electrodes are truly planes and not merely large segments of spheres of very large radius).

There is a well-known and accepted proof, depending on the theory of dimensions, that the current varies as the three-halves power of the voltage for any system of electrodes: this generalization encourages the hope that the argument for cylindrical, spherical and planar electrodes is also general and not true only for very special cases. If the anode is a closed metal chamber, its inside surface must be an equipotential surface and therefore the electric force must be normal to it. Moreover, since it is a metal, the electric force must be zero everywhere within its thickness and consequently must be zero also everywhere outside the chamber.

It is now appropriate to remind the reader of Green's theorem, which states that 'if a closed equipotential surface (i.e. a surface in the purely mathematical sense and not only a metal chamber) surrounds any number of charges, the force at any point outside this equipotential is the same as if the enclosed charges were removed and replaced by a charge density equal to $-\frac{1}{4\pi} \frac{\partial V}{\partial n}$ on the closed surface'. There is a second portion of Green's theorem which never seems to be developed explicitly in the classic textbooks, although this can be done by extending the mathematical analysis used in proving the first part. It is that, if the electric density $-\frac{1}{4\pi} \frac{\partial V}{\partial n}$ really did exist over the closed equipotential, it is a consequence of the inverse square law that the electric force would be zero at all points within the closed equipotential surface. The general mathematical analysis shows that this is necessarily true, although consideration will show that it is a very surprising result. Since the electric force must be zero everywhere outside the anode, because it is a metal, the first part of Green's theorem shows that the positive charge on the inside of the anode must be distributed with a density $-\frac{1}{4\pi} \frac{\partial V}{\partial n}$, where V is to be calculated only from the space charge in transit between barrier and anode. The second part of Green's theorem proves that the positive charge on the inside of the anode cannot produce any electric force at any point inside the anode, no matter what its shape. So at least this portion of the present treatment of the diode has been generalized, and, accordingly, the flight of electrons to it is in no way assisted by the charge on the anode itself, i.e. the flying electrons are never attracted by the anode.

Closed surfaces must exist between the barrier and the anode, at which all electrons have the same kinetic energy, although this velocity is not necessarily normal to it. Then, by definition, these surfaces are equipotentials: they have charges inside them and charges between them and the anode. It is only if the electron density (which depends on the electron trajectories) over these equipotentials happens to satisfy the condition $-\frac{1}{4\pi} \frac{\partial V}{\partial n}$ that the force inside them is zero. Only then will the electric force on any electron be due only to the total charge inside the equipotential in which it is situated at a given moment. The author is unable to generalize the condition that 'the force on every electron arises only from the electrons which are behind it'. He has shown that this is true for every electron as it reaches the anode, and it does seem nearly true at all points of the path—indeed it is true in the spherical, cylindrical and planar cases.

(3) ELECTRIC FORCE INSIDE A SQUIRREL-CAGE GRID

Now consider a grid, squirrel cage in form as shown in Fig. 3 and bearing a negative charge q per unit length of each wire, the cage having a radius R less than a and greater than b . It is

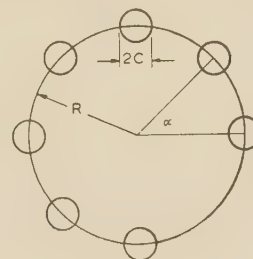


Fig. 3

advantageous first to consider the system when the cathode is not emitting electrons.

Let a closed Gaussian surface be formed within the squirrel cage. Then, since the charge is external to this surface, the integral of the normal component of electric force all over this surface is necessarily zero. But it will not be zero at every point of the surface: if it is outwards at one point it will be inwards at some nearby point.

Straightforward calculation shows that at radius r , for N wires,

$$E_r = \frac{2qN}{r} \sum_1^{\infty} \left(\frac{r}{R}\right)^{nN} \cos nN\theta, \text{ for } r < R \quad (11)$$

$$\text{or} \quad = \frac{2qN}{r} \left[1 + \sum_1^{\infty} \left(\frac{R}{r}\right)^{nN} \cos n\theta \right] \text{ for } r > R \quad (11a)$$

$$\text{If } N = 6 \text{ and } \frac{r}{R} = \frac{1}{10}, \text{ then } E_r \simeq \frac{2q}{R} \times \frac{1}{10^5}$$

This shows that the radial forces inside the cage are very small compared with the force at the outside surface of a tube of radius R , so long as $R - r$ is large compared with the radius of a wire of the cage.

If the charge on the wires of the cage is negative, no electrons emitted by the cathode (now supposed to be hot) will be collected by them. But the comparatively small tangential forces due to the wires will deflect the electrons slightly and cause some concentration of the flow between the wires.

If the radius b of the barrier is much smaller than R , the forces at it, due to the charged wires, will be very small. Doubtless they will produce a corrugation of the barrier surface, but of very small amplitude.

It seems reasonable to suppose that the charge on the grid will not, *per se*, have any appreciable effect on the barrier, and that it may be assumed, to a first approximation, that the current crossing the barrier is not determined by the grid *per se*. (At first sight this statement may well be so surprising as to seem absurd, but it is hoped that further reading will remove the difficulty.) Let the matter be put this way: Let the positive charge on the cathode be held constant by some means. Then it seems reasonable to expect that the current crossing the barrier, from a cathode of given temperature, will scarcely be affected by the charge on the grid; and furthermore, that the passage of this given current up to the cage radius R will scarcely depend on the existence of the grid. But once an electron is just outside the cage radius it will experience a repulsive force which scarcely differs from what would obtain if the grid were a continuous surface; this repulsive force from the negative grid will accelerate it towards the anode and cause it to reach the anode with a larger velocity than it would have if the given current were a

diode stream, with the grid absent. This means that, if a given current is passing to the anode of a given diode, which contains also a negative grid, the velocity of an electron at the anode (i.e. the anode potential) will be higher than if the grid were absent: this is in agreement with experience.

The idea that a negative grid accelerates the electrons passing to the anode may seem rather disturbing but it is true: in current jargon a negative grid is commonly said to brake the electrons, but this idea is false.

We are not saying that a negative grid does not reduce the current for a given anode voltage (see Section 13.1), but that for a given current the anode voltage must be increased when the grid is negative. The reader is asked to remember that the basic equation (1), namely $I_a = f(V_a + \mu V_g)$, is referring to a given current and not to the form of the function itself.

(4) MOTION INSIDE A PLANAR TRIODE

Now consider Fig. 4, which is meant to represent a planar triode with a grid distant g from the barrier and d from the anode. The charge on the grid is negative and of amount Q per unit area

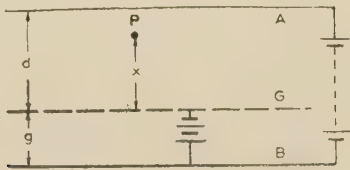


Fig. 4

of electrode plane. Although the algebra of a planar system will be used, it is probably wiser to regard the electrodes as being cylindrical but such that d and g are very small compared with the radius of the anode, the radius of the cathode as being comparable with that of the anode, the emission as being so very copious that g is not greater than d , and the radius of the wires of the grid as being small compared with g . All these conditions are realizable in principle. Thus it is possible to adhere to cylindrical electrodes, which are realizable in practice, and yet to use the algebra of a planar system as a very close approximation.

Let a given current, I , per unit area leave the barrier plane with negligible initial velocity. Following the argument set out, with reason, in the previous Section we shall suppose that this current is not disturbed appreciably by the negative charge on the grid. Or we can be more modest in this condition and suppose that, if it is necessary to make a very small change in the cathode emission in order to maintain I constant, there is no appreciable change in the distance g . In short, we suppose that, if the negative charge on the grid is changed and the anode voltage is adjusted to bring I back to its previous value, there will be no appreciable change in the motion of electrons between B and G so long as they are not actually above the plane through the axes of the grid wires. Thus, provided that any necessary adjustments have been made to maintain I constant, the motion of electrons up to the grid plane is sensibly the same as in a diode of spacing g carrying a current I and thus having a velocity u_0 in the grid plane corresponding to a potential V_0 such that

$$I = \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} \frac{V_0^{3/2}}{g^2}$$

The electrons gain kinetic energy between the plane of the grid and the anode by reason of the unresisted expansion of the electron gas between these two planes. Let V' be the potential which the anode would have if the grid were not present. Between the anode and the plane of the grid wires the negative

charge, Q , on the grid (together with its positive counterpart on the anode) produces a uniform force $4\pi Q$ directed towards the anode. This constant force will do work $V'' = 4\pi Qd$ on the electrons and increase their kinetic energy during their passage from the grid plane to the anode. If the very reasonable assumption is made that the imparting of this additional kinetic energy does not change the gain of kinetic energy which arises from the unresisted expansion of the electron gas, the net velocity at the anode will correspond to a potential V_a such that

$$V_a = 4\pi Qd + V' \quad (12)$$

In fact, however, very close analysis shows that the relationship given by eqn. (12) is not exact, because there is a very slight diminution of the work done by the unresisted expansion. This analysis is set out in detail in Section 13.2, where it is shown that the lack of precision in eqn. (12) is entirely negligible. Accordingly, with this conscious caveat, eqn. (12) will be maintained and treated as though it were exact.

Let t_1 be the time for an electron to traverse the distance g between barrier and grid plane when there is an assigned current I leaving the barrier: then the total charge in transit between barrier and grid is It_1 .

The charge on the grid, in a valve of given construction, depends only on the grid potential; it is an independent variable, so let its amount be Q . When the current leaving the barrier has some assigned value I , the charge behind the grid is It_1 . It is convenient to express Q in terms of It_1 by the equation $Q = yIt_1$, where y is a non-dimensional factor. Of course, it will turn out eventually that I is a function of V_a , but, strictly speaking, Q and I are independent in that Q can be given any desired value by suitable adjustment of the grid potential only, and then I can be brought to any desired value by adjustment of anode potential. Thus the factor y can be brought to any desired value and has all the characteristics of an independent variable.

Accordingly, eqn. (12) will now be written in the form

$$V_a = 4\pi yIt_1d + V' \quad (13)$$

From eqn. (10c)

$$V_0 = 3\pi It_1g$$

whence

$$V_a = \frac{4}{3}y \frac{d}{g} V_0 + V' = \frac{4}{3}yz V_0 + V' \quad (14)$$

where $z = d/g$

From eqn. (10b)

$$\frac{V'}{V_0} = \left(\frac{d+g}{g} \right)^{4/3} = (1+z)^{4/3}$$

whence

$$\frac{V_a}{V'} = 1 + \frac{4}{3} \frac{z}{(1+z)^{4/3}} y \quad (15)$$

This shows that for any assigned anode current there is a precisely linear relationship between the anode voltage and the charge on the grid. It now remains only to show that for any assigned anode current the charge on the grid is proportional to the negative voltage of the grid and then a relationship of the form $(V_a + \mu V_g) = \text{constant}$ will have been established. Note that the effect of space charge will not have been ignored, since it is the space charge alone which propels the current. The classic derivations of the relationship $(V_a + \mu V_g) = \text{constant}$ were evolved on the assumption that there was no space charge. Here the problem of a real triode will have been simplified only by the very reasonable approximation that, for a given current, the flow is not modified appreciably by the presence of the grid.

It is worth noting that the parameter $\frac{4z}{3(1+z)^{4/3}}$ changes

very slowly with z , passing through the maximum value 0.63 when $z = 3$; thus when $z = 1$ its value is 0.53 and when $z = 7$ its value is 0.584. Accordingly, the slope of the line relating V_a/V' and y , eqn. (15), is virtually independent of d/g in the range of this ratio likely to be met with in practice.

(5) RELATION BETWEEN THE POTENTIAL OF THE GRID AND THE CHARGE ON IT WHEN THERE IS A GIVEN CURRENT TO THE ANODE

Consider an infinite grating of round wires, each of radius c , separated by a distance S centre to centre. Let each wire carry a charge q per unit length; the charge is supposed to be distributed sensibly uniformly round the wires, and this requires that $c/S \ll 1$.

Then it is well known^{7,8} that the force at distance x (along the normal) from the mid-plane of the grating is given by

$$E = \frac{2\pi q}{S} \coth \frac{\pi x}{S} \quad (16)$$

$$E = \frac{2\pi q}{S} \tanh \frac{\pi x}{S} \quad (17)$$

according as the normal passes through a wire or through the point midway between two wires respectively.

The positive charge on the anode of a planar diode must be equal numerically to the total charge on the grid wires plus the total electricity in transit between the barrier and the anode. Accordingly, the net force between grid and barrier which results from the presence of charge on the grid (together with the equal positive charge on the anode) is, at a point on a normal through a wire,

$$E' = \frac{2\pi q}{S} \left(\coth \frac{\pi x}{S} - 1 \right) = \frac{2\pi q}{S} \times 0.036 \text{ when } x = \frac{2}{\pi} S$$

$$\text{or } \frac{2\pi q}{S} \times 0.0037 \text{ when } x = S$$

Thus the net force has fallen to a very small value at distances greater than about S . This means that, provided g is several times S , the force at points between the grid and the barrier depends almost entirely on the space charge. Or, in the words used previously, in this region the flight of electrons, carrying a given current, will scarcely be dependent upon the presence of the grid. The difference of potential between the surface of a grid wire and the barrier, arising from the negative charge placed on the grid by means of the battery connected to it, is

$$V_1 = \int_c^g E' dx = \frac{2\pi q}{S} \int_c^g \left(\coth \frac{\pi x}{S} - 1 \right) dx$$

$$= \frac{2\pi q}{S} \left[\frac{S}{\pi} \log_e \sinh \frac{\pi x}{S} - x \right]_c^g$$

$$= \frac{2\pi q}{S} \left[\frac{S}{\pi} \log_e \sinh \frac{\pi g}{S} - \frac{S}{\pi} \log_e \sinh \frac{\pi c}{S} - g + c \right]$$

$$\equiv qB$$

where B is a purely geometrical factor. If g/S is large then

$$B \simeq \frac{2\pi}{S} \left[\frac{S}{\pi} \log_e \frac{e^{\pi c/S}}{2} - \frac{S}{\pi} \log_e \sinh \frac{\pi c}{S} - g + c \right]$$

$$= 2 \left(\log_e 2 \sinh \frac{\pi c}{S} - \frac{\pi c}{S} \right)$$

Now V_1 , the work done in taking unit charge from the barrier plane to a grid wire against the force due to the negative charge on the grid wires and the corresponding positive charge on the anode, is proportional to the shaded area ABCD in Fig. 5. For

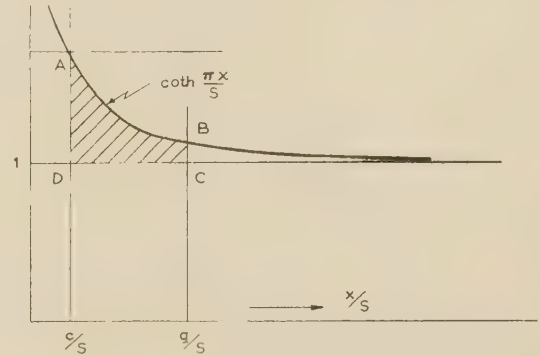


Fig. 5

a given value of c/S , i.e. a given grid, it is obvious from this Figure that the shaded area increases as g/S increases, up to a limiting value which is given by the approximate value of B given above. If the cathode emission is increased, for a given anode current, then the barrier plane will approach the grid plane, which means an appropriate decrease of g . Note, however, that, provided only that the grid wires can be treated as filaments, which means c/S is small, B can be calculated correctly for any value of g/S and there is no other simplification or approximation in the value of V_1 . The equipotentials round a grating of filamentary wires can readily be plotted and are shown, for example, in Fig. 137 of Reference 8, in which they begin as symmetrical ovals which differ very little from a circle. The potential of the grating, referred to a filament as origin is given by

$$V = q \log_e \left(\cosh \frac{2\pi x}{S} - \cos \frac{2\pi y}{S} \right)$$

at a point distant x from the plane of the filaments and distant y from the normal plane through a filament. This equation gives the ratio between the major and minor axes of the oval equipotentials by equating the potential at the point $(x, 0)$ to that at the point $(0, y)$, from which is obtained

$$\cos \frac{2\pi y}{S} = 2 - \cosh \frac{2\pi x}{S}$$

Some solutions of it are given in Table 2.

Table 2

x/S	y/S	y/x
0.1	0.1039	1.039
0.2	0.235	1.17
0.25	0.333	4/3
0.28	0.500	1.76

This shows that even if y/S is as large as 0.1039, the equipotential is oval only to the extent of the ratio being 1.039. Although it is impracticable to solve correctly a grating of truly circular wires of finite diameter, it has now emerged that the solution for a grating of filaments is still exact for wires of appreciable size, provided only that they are not truly round, but are oval only to an extent which might well be within the limits of manufacture of ostensibly round wires. Thus it appears that the expression for B must be valid for all values of g/S and for values of c/S up to at least 0.1.

The work done in taking unit charge from the barrier is V_1

less the work done against the forces arising from the electrons which are in transit between the barrier and the plane of the grid; this has been called V_0 ; let the difference be V_g ; then $V_g = qB - V_0$.

Now V_0 will depend on the initial velocity of emission from the barrier and also on the extent to which the density varies across a given plane as the result of disturbance produced by the existence of the grid. If this variation of density is ignored (i.e. g/S is large) and the emission velocity is zero, then $V_0 = 3\pi(It_1)g$, by eqn. (10a). Then $V_g = QSB - 3\pi(It_1)g$, where Q is the charge per unit area of grid plane, i.e.

$$V_g = y(It_1)SB - 3\pi(It_1)g$$

Therefore
$$y = \frac{3\pi g}{B S} + \frac{V_g}{(It_1)SB} \quad (18)$$

Accordingly, y is a linear function of V_g . It has been shown in eqn. (15) that y is a linear function of V_a and thus V_a is a linear function of V_g for any given value of I . Now $V_a/V_0 = (1+z)^{4/3} + \frac{4}{3}yz$, from eqns. (14) and (15).

Therefore
$$y = \frac{3g}{4d} \frac{V_a}{V_0} - \frac{3}{4} \frac{(1+z)^{4/3}}{z}$$

and
$$y = \frac{3\pi g}{B S} + \frac{V_g}{(It_1)SB} \quad (19)$$

whence
$$\frac{1}{4\pi} \frac{1}{d} \frac{V_a}{(It_1)} - \frac{3}{4} \frac{(1+z)^{4/3}}{z} = \frac{3\pi g}{B S} + \frac{V_g}{(It_1)SB}$$

since
$$V_0 = 3\pi(It_1)g$$

Therefore
$$\frac{1}{4\pi d(It_1)} \left(V_a - \frac{4\pi d}{SB} V_g \right) = \frac{3}{4} \frac{(1+z)^{4/3}}{z} + \frac{3\pi g}{B S}$$

Hence, if It_1 is constant, as it is for any assigned value of I , then

$$V_a - \frac{4\pi d}{SB} V_g \text{ is constant}$$

i.e.
$$V_a + \frac{4\pi d V_g}{2S \left(\log_e 2 \sinh \frac{\pi c}{S} - \frac{\pi}{c} \right)} \text{ is constant}$$

i.e.
$$V_a - \mu V_g \text{ is constant,}$$

where
$$\mu = - \frac{2\pi d}{S \log_e 2 \sinh \frac{\pi c}{S}} \quad (20)$$

if $\pi c/S$ is ignored.

Thus the well-known and classic expression for μ has been obtained by a method which takes full account of the existence of space charge.

After some reduction, these expressions give

$$I = \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} \frac{1}{(g+d)^2} \frac{(V_a + \mu V_g)^{3/2}}{\left[1 + \frac{S}{g} \frac{\mu}{(1+z)^{4/3}} \right]^{3/2}} \quad (21)$$

(6) THE CYLINDRICAL TRIODE

In Section 2 the equation for a cylindrical diode was given in the form

$$I = \frac{1}{9\pi} \left(\frac{2e}{m} \right)^{1/2} D^2 \frac{V^{3/2}}{g^2}$$

where the parameter D^2 was tabulated in Table 1.

Using the previous notation, let V' be the potential of the anode of a diode at a distance $(g+d) = g(1+z)$ from the barrier, which is passing an assigned current I , and let V_0 be the potential of the anode of a diode at a distance g from a similar barrier, which is also passing the same assigned current I . Then for this same current in each diode

$$D_0^2 V_0^{3/2} = \frac{D_1^2 V'^{3/2}}{(1+z)^2}$$

It follows from Table 1 that D_0^2 will be very nearly equal to D_1^2 provided that $(g+b)/b$ is not less than about 3. In such circumstances $V'/V_0 = (1+z)^{4/3}$, as it did with planar electrodes. Even if $(g+b)/b \approx 1.3$, then $V'/V_0 \approx 1.1(1+z)^{4/3}$; accordingly, it is substantially correct to write $V'/V_0 = (1+z)^{4/3}$ except in circumstances when the barrier is almost coincident with the grid—and these circumstances are really excluded by the condition that the clearance between the grid and the barrier must be large compared with the pitch between the wires of the squirrel-cage grid. The equation corresponding to eqn. (13)

now becomes $V_a = 2Q \log_e \frac{a}{R} + V'$, where R is the radius of the grid. As before, let t_1 be the time of flight from the barrier radius to the grid radius; then the total charge in transit is It_1 .

Let $Q = yIt_1$; then

$$V_a = 2yIt_1 \log_e \frac{a}{R} + V'$$

It is a simple matter to show⁹ that

$$V_0 = \frac{3}{2} It_1 \frac{\beta}{\alpha}$$

where
$$\beta = y \left(1 - \frac{2}{5} \rho + \frac{11}{120} \rho^2 - \dots \right)$$

and
$$\alpha = \left(1 - \frac{3}{10} \rho + \frac{9}{120} \rho^2 - \frac{49}{4400} \rho^3 + \dots \right)$$

and
$$\rho = \log_e \frac{R}{b}$$

Whence
$$V_a = \frac{4}{3} \frac{\alpha}{\beta} V_0 \log_e \frac{a}{R} + V'$$

or
$$\frac{V_a}{V'} = 1 + \frac{4}{3} \frac{1}{(1+z)^{4/3}} \frac{\alpha}{\beta} \log_e \frac{a}{R} = 1 + Fy \dots$$

Some values of α/β are collected in Table 3

Table 3

R/b	4.0	3.0	2.0	1.50
α/β	0.85	1.00	1.60	2.50

It is instructive to assess the numerical value of F , the coefficient of y in the equation above; this will be done for $a/R = 2$. Some results are collected in Table 4.

Table 4

R/b	1.5	1.65	2	3	4
z	4	2.54	2	3/2	4/3
F	0.27	0.34	0.34	0.27	0.25

This shows that F remains reasonably constant over a considerable range of R/b ; in a given structure with $a/R = 2$, an increase of R/b corresponds to a diminution of cathode temperature for a given anode current. For the limiting value of F see Section 13.2.

(7) RELATION BETWEEN THE POTENTIAL OF A SQUIRREL-CAGE GRID AND THE CHARGE ON IT WHEN THERE IS A GIVEN CURRENT TO THE ANODE

Integration of eqn. (11) gives

$$-V = 2q \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{r}{R}\right)^{nN} \cos nN\theta$$

or $V = 2q \log \left[1 - \left(\frac{r}{R}\right)^N \right]$, when $\theta = 0$

Hence the potential difference, V_1 , between the surface of a grid wire and the barrier, due to q , will be

$$\begin{aligned} V_1 &= 2q \log \frac{1 - \left(1 - \frac{c}{R}\right)^N}{1 - \left(\frac{b}{R}\right)^N} \\ &\simeq \frac{2Q}{N} \left\{ \log_e \frac{Nc}{R} - \log_e \left[1 - \left(\frac{b}{R}\right)^N \right] \right\} \\ &= -\frac{2Q}{N} \left\{ \log_e \frac{R}{Nc} + \log_e \left[1 - \left(\frac{b}{R}\right)^N \right] \right\} \equiv QB' \dots \end{aligned}$$

Whence $V_g = QB' - V_0$

$$= y(I t_1) B' - \frac{3}{2} (I t_1) \frac{\beta}{\alpha}$$

Therefore $y = \frac{3}{2} \frac{\beta}{\alpha} \frac{1}{B'} + \frac{V_g}{(I t_1) B'}$

From this it follows that

$$\begin{aligned} \frac{1}{(I t_1) 2 \log_e \frac{a}{R}} \left[V_a - \frac{2 \log_e \frac{a}{R}}{B'} V_g \right] &= \frac{3}{2} \frac{\beta}{\alpha \beta'} + \frac{1}{F} \\ &= \frac{3}{2} \frac{\beta}{\alpha} \left[\frac{1}{B'} + \frac{1}{2} \frac{(1+z)^{4/3}}{\log_e \frac{a}{R}} \right] \end{aligned}$$

In this case

$$\mu = \frac{\log_e \frac{a}{R}}{\frac{1}{N} \left\{ \log_e \frac{R}{Nc} + \log_e \left[1 - \left(\frac{b}{R}\right)^N \right] \right\}} \simeq \frac{N \log_e \frac{a}{R}}{\log_e \frac{R}{Nc}} \quad (22)$$

which is the well-known expression derived long ago for the case when space charge is absent.

It has now been shown that, for a given anode current, $V_a + \mu V_g$ is precisely constant, where μ is the parameter which was deduced long ago from a proof which ignored space charge.

After a little reduction the foregoing equations give

$$I = \frac{2}{9} \left(\frac{2e}{m} \right)^{1/2} \frac{(V_a - \mu V_g)^{3/2}}{[\mu + (1+z)^{4/3}]^{3/2}} \frac{1}{R \beta^2} \quad (23)$$

This agrees with the expression given by Eccles¹⁰ if $\beta^2 = 1$ and $z = 0$, and with that given by Appleton¹¹ if $(1+z)^{4/3}$ is ignored in comparison with μ .

In eqn. (23) β^2 refers to the ratio R/b ; the formula would be more satisfactory if it involved a rather than R , but this can scarcely be done because of the parameter β^2 . In it $z = (a - R)/(R - b)$ and thus, in general, is likely to be larger than

unity, since b is the radius of the barrier: accordingly $(1+z)^{4/3}$ may well be comparable with μ .

For a given cathode temperature b will decrease as I increases; this will involve a change of β^2 and z , and thus I will not vary accurately with $(V_a - \mu V_g)^{3/2}$. For any given current the plot of V_a and V_g will be a straight line, but the lines for different current will be parallel only in so far as μ can be regarded as independent of b . The denominator of the expression for μ is equal to

$$\left\{ \log_e \frac{R}{Nc} + \log_e \left[1 - \left(\frac{b}{R}\right)^N \right] \right\} \simeq \log_e \frac{R}{Nc} - \frac{Nb}{R}$$

Accordingly, as b decreases the denominator increases, which means that μ will decrease slightly as I increases, in accordance with familiar experience; or, for a given current, μ tends to increase with an increase of cathode temperature.

(8) THE SMALLEST VALUE OF y FOR WHICH THE GRID CURRENT CAN BE ZERO

Throughout this discussion the charge maintained on the grid, by means of the battery connected between grid and cathode, has been taken as negative. Moreover it has been taken to be sufficiently negative to prevent any electrons reaching the grid. It is of interest to discover the smallest value of y for which this condition is fulfilled.

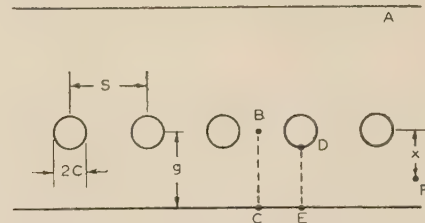


Fig. 6

Consider a grid typified by Fig. 6 and let it carry a negative charge of $-q$ per unit length of each wire. Then at points B the electric force due to this charge is zero and at points D it is nearly equal to $2q/c$. Let the parallel anode A carry a positive charge q/S per unit area. Then, from eqn. (17), the electric force E_1 at a point on BC is

$$E_1 = \frac{2\pi q}{S} \left(1 - \tanh \frac{\pi x}{S} \right)$$

towards the anode.

At a point on DE the electric force E_2 is, by eqn. (16), given by

$$E_2 = \frac{2\pi q}{S} \left(\coth \frac{\pi x}{S} - 1 \right)$$

away from the anode.

The work done along these two paths is respectively

$$V_1 = 2q \log 2$$

and

$$V_2 = 2q \left(\log_e 2 \sinh \frac{\pi c}{S} - \frac{\pi c}{S} \right) \equiv qB$$

Let a current I per unit area be emitted from the barrier. Along CB the electrons are arrested in their flight by the presence of the charge on the grid and on the anode. Along ED their flight is impeded by the presence of this charge. At the point D the potential is

$$V_g = qB - V_0$$

where

$$V_0 = 3\pi I t_1 g$$

The limiting condition for an electron to reach a grid wire is that $V_g = 0$ and then

$$V_0 = qB \\ = QSB$$

Therefore

$$3\pi It_1 g = yIt_1 BS$$

so that

$$y = \frac{3\pi g}{BS}$$

$$\frac{3}{4} \frac{\mu}{z}, \text{ since } \mu \equiv \frac{4\pi d}{BS}$$

Accordingly, the smallest value of y is likely to be of the order of $\frac{1}{2}\mu$ —say not less than 5 in a normal triode.

From eqn. (15), the least value of V_a/V' is

$$\left[1 + \frac{\mu}{(1+z)^{4/3}} \right]$$

the numerical value of which is not likely to be less than about 5. At the point B the potential is

$$V'_g = V_0 + 2q \log_e 2 \\ = V_0 + 2QS \log_e 2 \\ = 3\pi It_1 g + 2yIt_1 S \log_e 2 \\ = V_0 \left(1 + \frac{2yS \log_e 2}{3\pi g} \right)$$

(9) COMPARISON OF THE PRESENT METHOD WITH THE OLDER ONES

As already stated, it was noticed by Van der Bijl in 1913¹ that $I_a = f(V_a + \mu V_g)$ and his book,² published in 1920, implies that his theoretical deduction of this property was based on the absence of space charge. But we shall ignore this at present and start by outlining the method used by Eccles¹⁰ in 1921. On p. 333 of this book he develops the theory of the 'triode with plane electrodes' and says

It is possible to attack the problem by a method based upon the calculation of capacity coefficients, making allowance afterwards for the influence of the space-charge upon the electric field between the electrodes . . . but it is more instructive to make a direct attempt upon the problem.

He assumes that the electric field is practically constant between the anode and a point B in the plane of the grid wires but midway between any pair; then, if σ_a is the charge per unit area of anode, $V_a - V_B = 4\pi\sigma_a d$. If the charge on the grid wires, per unit area of the grid plane is Q then he states that $\sigma_a = Q$ + all space charge: in this we agree. He then states that $\sigma_a = \alpha Q$,

where α is a quantity to be determined by experiment and obviously approaches unity when the space charge, that is to say, the electron current, is very small. Experiment shows that for most types of triode α is nearly equal to unity for a fair range of values of grid voltages and anode current, but the mode of derivation shows that it should tend to increase as the electron current increases.

Surely it is pertinent to inquire how it is possible to measure either σ_a or Q and thereby to determine that α is nearly equal to unity. Surely the equation $\sigma_a = \alpha Q$ is little more than a guess, which is, in fact, a wrong guess and is verging on neglect of space charge.

He then argues that the voltage difference between the point B and a point of a grid wire is of the form $V_B - V_g = \frac{2Q}{s} \log \frac{\beta}{c}$, 'where β is a geometrical constant that can be determined by

calculation for any number of wires if required'. With this we are in agreement, especially if space charge is ignored: the value of his constant β could virtually be derived from eqns (19a) and (19b). He then evaluates the voltage V_0 between the point B and the barrier and gives it as $V_B = (J/A)^{2/3} g^{4/3}$. This is precisely what we have done and is equivalent to the assumption that the passage of a given current is undisturbed by the charge on the grid, up to a point in the grid plane.

He then derives $V_a - V_B = 4\mu\alpha Qd$

$$= \frac{2\pi\alpha sd}{\log_e \beta/c} (V_B - V_g)$$

or

$$V_a + \frac{2\pi\alpha sd}{\log_e \beta/c} V_g = \left(1 + \frac{2\pi\alpha sd}{\log_e \beta/c} \right) V_B \\ = \left(1 + \frac{2\pi\alpha sd}{\log_e \beta/c} \right) \left(\frac{J}{A} \right)^{2/3} g^{4/3}$$

and thereby a relationship of the form $(V_a + \mu' V_g)$ is obtained. But it will be noticed that μ' contains the undetermined parameter α , said to be 'not far from unity'.

The solution seems to be sound in so far as V_B is regarded as being independent of the presence of the grid, although it is not explained why the presence of the grid scarcely disturbs the field between the grid and the barrier. Furthermore, the determination of V_g is sound; but it seems that in the region between grid and anode the existence of space charge, both between the grid and the anode and between the grid and the barrier, is virtually ignored.

We will now outline the treatment given in 1931 by Sir Edward Appleton.¹¹ He compares the triode with an 'equivalent diode' whose electrode separation is g and says:

We require to find the potential, v , at which the 'anode' of the equivalent diode must be maintained in order that its effect may simulate the combined effect of the anode and grid of the triode which are maintained at potentials v_a and v_g respectively. In doing so we shall as a first approximation neglect the effect of the space charge. . . . We require to find the value of v which will make the charge per unit area on (4) [the equivalent diode] equivalent to the combined charge on (2) [the grid] and (3) [the anode of the triode] . . . It is evident that the charges per unit area of these two condensers will be equivalent when $C_{12}v_g + C_{13}v_a = C_{14}v$, where the C 's represent capacity coefficients.

$$\text{Whence} \quad v = \frac{C_{12}}{C_{14}} (v_g + \frac{C_{13}}{C_{12}} v_a)$$

and the coefficient of v_a is the same as the parameter we have already called μ . Now $C_{12}v_g + C_{13}v_a$ represents the charge per unit area on the cathode plane of the triode in the absence of space charge: it is far from easy to see how this is relevant to the triode problem. In the corresponding treatment, Van der Bijl¹ (p. 228) obtains an expression for the charge density, σ , on the cathode plane in the presence of the grid and anode at assigned voltages and says:

The quantity σ gives a measure of the intensity of the field near [the cathode], which affects the flow of electrons from it. These equations apply, of course, to the case in which there are no electrons in the space between the electrodes to cause a distortion of the field. As far as the determination of the screening effect of the grid for most practical purposes is concerned, it is found that the presence of the electrons in the space usually does not materially influence the results.

Reading this more than 30 years after it was written, it is now difficult to understand what is meant by the idea that 'the field near the cathode affects the flow of electrons from it'. And difficult to understand how it could be discovered that the

presence of electrons in the space usually does not materially influence the screening effect of the grid.

Van der Bijl and his contemporaries obtained the right formula, but the thoughts which led them to do so seem now to be obscure. We have shown that their formula was more correct than they supposed it to be.

Appleton (*loc. cit.*, p. 41) says

In a more detailed investigation, in which the effect of space charge is taken into account, Schottky¹² has shown that instead of

$$\text{the equation} \quad v = \frac{v_g + Dv_a}{1 + D \frac{\log b/\kappa}{\log a/\kappa}}$$

$$\text{where} \quad D \equiv \frac{1}{\mu}$$

$$\text{we have} \quad v = \frac{v_g + Dv_a}{1 + D \left(1 + \frac{2}{3} \log \frac{b}{a} \right)}$$

Since, however, in most practical cases, D is small compared with unity, it is sufficiently accurate, even when the space charge is present, to write $v = v_g + Dv_a$.

This excerpt is given only to emphasize that the early writers were fully conscious that the effect of space charge had been omitted.

Chaffee,¹³ writing in 1933, follows substantially the treatment of Van der Bijl and does not attempt any further serious treatment of the effect of space charge.

A completely revised and rewritten sixth edition¹⁴ of Appleton's book, in collaboration with W. H. Aldous, was first published in 1952. On pp. 32-35 of it a much more serious attempt is made to discuss the effect of space charge. But it proceeds by means of the capacitance-coefficients method and not by the approach used in the present paper. It arrives at a formula very similar to that given by Schottky in 1920.

(10) ACKNOWLEDGMENT

The author wishes to tender his very grateful thanks to Mr. G. W. Warren, and others, of the Research Laboratories of The General Electric Company, Ltd., Wembley, for the great help given to him during the course of producing the paper, for close and detailed discussion of the validity of the method of approach which has been used and for making valuable suggestions about the order and method of presenting it.

(11) REFERENCES

- (1) VAN DER BIJL: 'Initial Energies of Photoelectrically-Liberated Electrons', *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 1913, **15**, p. 330.
- (2) VAN DER BIJL: 'The Thermionic Vacuum Tube and its Applications' (McGraw-Hill, New York, 1920).
- (3) MOULLIN, E. B.: 'The Three Halves Power Law of the Diode', *Wireless Engineer*, 1937, **14**, p. 193.
- (4) MÖLLER, H. G.: 'Die Physikalischen Grundlagen der Hochfrequenztechnik', p. 35.
- (5) LANGMUIR, I., and BLODGETT, K. B.: 'Currents Limited by Space Charge between Coaxial Cylinders', *Physical Review*, 1923, **22**, p. 347.
- (6) MOULLIN, E. B.: 'Spontaneous Fluctuations of Voltage' (Clarendon Press, Oxford, 1938), p. 80.
- (7) MAXWELL, J. C.: 'A Treatise on Electricity and Magnetism' (Clarendon Press, Oxford, 1873), Vol. I, p. 286.
- (8) MOULLIN, E. B.: 'The Principles of Electromagnetism' (Clarendon Press, Oxford, Third Edition, 1955), p. 225.
- (9) MOULLIN, E. B.: 'Time Flight of Electrons', *Wireless Engineer*, 1935, **12**, p. 371.

- (10) ECCLES, W. H.: 'Continuous-Wave Wireless Telegraphy' (Wireless Press, London, 1921), p. 337.
- (11) APPLETON, E. V.: 'Thermionic Vacuum Tubes' (Methuen, London, 1931), p. 41.
- (12) SCHOTTKY, W.: 'Zur Raumladungstheorie der Verstärkeröhren', *Wissenschaftliche Veröffentlichungen aus dem Siemens-Konzern*, 1920-22, **1**, p. 64.
- (13) CHAFFEE, E. L.: 'Theory of Thermionic Vacuum Tubes' (McGraw-Hill, New York, 1933).
- (14) ALDOUS, W. H., and APPLETON, E. V.: 'Thermionic Vacuum Tubes and their Applications' (Methuen, London, 1952).

(12) APPENDIX

(12.1) The Means by which V_g Controls I when V_a is Constant

One of the most familiar experiences with a triode is the decrease of anode current when the grid is made more negative, the anode voltage being kept constant: how does this accord with our approach? If there is an increase of negative charge on the grid, there is an increase of the energy imparted to the electrons as they pass from the plane of the grid to the anode. But, because the anode voltage is constant, the electrons must continue to arrive at the anode with an unchanged velocity, notwithstanding the additional negative charge on the grid. Therefore the additional negative charge on the grid must somehow cause them to arrive at the grid plane with a decreased velocity.

From the equation $x = \frac{t}{3}(u + u_0)$ it follows that, if x is constant, t must increase if u decreases. And from the equation $t_1 = \left(\frac{6g}{AI}\right)^{1/3}$ it follows that if t_1 increases then I must decrease, hence proving that a decrease of current must result from an increase of negative charge on the grid.

If the current decreases, the cathode temperature being constant, there must be an increment of positive charge on the cathode. If more negative charge is given to the grid there must necessarily be an increase of positive charge on the anode: this, it seems, has to be accompanied by an appropriate increase of positive charge on the cathode, but the mechanism by which this must necessarily occur is not clear. This problem is not peculiar to a triode, but arises equally in a diode: it involves the stability of the barrier surface, and that is a problem which does not appear to have been tackled.

(12.2) Evaluation of V_a from the Electron Velocity at the Anode

For an electron at distance x above the plane of the grid the equation of motion is

$$\begin{aligned} \frac{d^2x}{dt^2} &= v \frac{dv}{dx} = A(Q + It_1 + Itx) \\ &= AI[(1 + y)t_1 + t] \quad \dots \quad (24) \end{aligned}$$

where x and t are the only variables.

$$\text{Therefore} \quad \left[\frac{v^2}{2} \right]_{v_1}^{v_2} = AI \left[(1 + y)t_1 d + \int_0^d t dx \right] \quad \dots \quad (25)$$

$$\text{also} \quad v = \frac{dx}{dt} = AI \left[(1 + y)t_1 t + \frac{1}{2}t^2 + \frac{v_1}{AI} \right]$$

from eqn. (24)

$$= AI \left[(1 + y)t_1 t + \frac{1}{2}t^2 + \frac{3g}{AI t_1} \right] \quad \dots \quad (26)$$

from eqn. (9).

Therefore

$$\left[\frac{v^2}{2} \right]_{v_1}^{v_2} = AI \left\{ (1+y)t_1 d + AI \int_0^t \left[(1+y)t_1 t^2 + \frac{1}{2} t^3 + \frac{3gt}{AI t_1} \right] dt \right\}$$

from eqns. (25) and (26)

$$\text{But } x = AI \left[(1+y)t_1 \frac{t^2}{2} + \frac{t^3}{6} + \frac{3gt}{AI t_1} \right] \quad (27)$$

from eqn. (26)

$$\text{or } (1+y)t_1 \frac{t'^2}{2} = \frac{d}{AI} - \frac{t'^3}{6} - \frac{3gt'}{AI t_1} \quad (28)$$

$$\text{Whence } \frac{v_2^2 - v_1^2}{2AI} = (1+y)t_1 d$$

$$+ AI \left[\frac{3t'}{2} \left(\frac{d}{AI} - \frac{t'^3}{6} - \frac{3gt'}{AI t_1} \right) + \frac{t'^4}{8} + \frac{3g}{2AI} \frac{t'^2}{t_1} \right]$$

from eqns. (26) and (28).

$$= t_1 d \left\{ (1+y) + \left[\frac{3}{2} \frac{t'}{t_1} \left(1 - \frac{2}{z} \frac{t'}{t_1} \right) - \frac{3}{4} \frac{1}{z} \frac{t'^4}{t_1^4} \right] \right\} \quad (29)$$

where $z \equiv d/g$

Consideration will show that eqn. (29) is very akin in form to eqn. (12), but for any given value of z , t' is a function of y , obviously decreasing as y increases. Accordingly eqn. (12) would be exact only if t' were not a function of y . It has been stated already, just after the derivation of eqn. (12), that it was inexact, but to a quite negligible extent. To demonstrate this negligibility it is necessary to evaluate numerically the magnitude of the term (in square brackets) involving t'/t_1 , for various values of y with successive values of z . Let the term involving t' be denoted by D . For the case when $z = 1$ some values of $(1 + D)$ for various values of y are shown collected in Table 5,

Table 5

$z = 1$

y	0	1	2	4	6	10	27.3	112
$(1 + D)$	1.185	1.185	1.183	1.163	1.155	1.149	1.12	1.067
t'/t_1	0.260	0.228	0.208	0.175	0.153	0.14	0.10	0.05

which shows that $(1 + D)$ is constant to within about $\pm 3\%$ up to $y = 25$, and becomes of progressively less importance compared with y . In other words, V_a is a linear function of y within extremely small limits.

For the case when $z = 7$ the corresponding values are given in Table 6.

Table 6

$z = 7$

y	0	1.14	2.1	3.6	6.2	10.9	24.5	49.6
$(1 + D)$	1.96	1.88	1.81	1.74	1.65	1.53	1.41	1.19
t'/t_1	1	0.8	0.7	0.6	0.5	0.4	0.3	0.2

This shows that, even in so extreme a case as $z = 7$, $(1 + D)$ is still constant within about $\pm 15\%$ up to $y = 25$.

It is shown in Section 8 that values of y smaller than about 6 do not arise. Eqn. (27) can be written in the form

$$z = \frac{3t'}{t_1} + 3(1+y) \frac{t'^2}{t_1^2} + \frac{t'^3}{t_1^3}$$

Tables 5 and 6 show that t'/t_1 is less than $\frac{1}{2}$ when y is greater than 7 for all practicable values of z : accordingly, the term in $(t'/t_1)^3$ is of little importance. If it is ignored, then

$$\frac{t'}{t_1} = \frac{\sqrt{\left(\frac{4yz + 4z + 3}{3} \right)} - 1}{2(1+y)}$$

$$\text{But } \frac{u_2}{u_1} = 1 + (1+y) \frac{2t'}{t_1} + \left(\frac{t'}{t_1} \right)^2$$

$$\text{whence } \frac{u_2}{u_1} = \sqrt{\left(\frac{4yz + 4z + 3}{3} \right)} + \left(\frac{t'}{t_1} \right)^2$$

$$\text{whence } \frac{V_a}{V_0} = \left(\frac{u_2}{u_1} \right)^2 \simeq \frac{4yz + 4z + 3}{3}$$

$$\text{or } \frac{V_a}{V'} = \frac{4z + 3}{3(1+z)^{4/3}} + \frac{4z}{3(1+z)^{4/3}} y \quad (15a)$$

Eqn. (15a) is a closer approximation than eqn. (15), in which the constant term is unity. If $z = 1$ the constant term in eqn. (15a) is equal to 0.92.

Because y is likely to be greater than 6 consideration in conjunction with Tables 5 and 6 shows that the term t in eqn. (24) is never important and accordingly the electrons move from grid to anode under the action of a force which is very nearly constant. This is equivalent to saying that the effect of space charge is negligible in the region between the grid and the anode, which justifies the first part of the assumption made by Eccles.¹⁰

(12.3) Limiting Value of F when R/b Tends to Unity

If R/b is near unity, $\rho = \log(R/b)$ is small and then

$$\frac{\alpha}{\beta} \simeq \frac{1 + \frac{1}{10}\rho}{\rho} = \frac{1 + \frac{1}{10} \log_e \left(1 + \frac{g}{b} \right)}{\log_e \left(1 + \frac{g}{b} \right)} \simeq \frac{b}{g} \left(1 + \frac{3g}{5b} \right)$$

$$\log_e \frac{a}{R} = \log_e \left(1 + \frac{d}{R} \right) \simeq \frac{d}{R} \left(1 - \frac{1}{2} \frac{d}{R} \right), \text{ if } \frac{d}{R} \ll 1$$

$$\text{Then } F = \frac{4}{3} \frac{1}{(1+z)^{4/3}} \frac{b}{g} \left(1 + \frac{3g}{5b} \right) \frac{d}{R} \left(1 - \frac{1}{2} \frac{d}{b+g} \right)$$

$$= \frac{4}{3} \frac{z}{(1+z)^{4/3}} \frac{\left(1 + \frac{3g}{5b} \right)}{\left(1 + \frac{g}{b} \right)} \left(1 - \frac{1}{2} \frac{d}{b+g} \right)$$

$$\simeq \frac{4}{3} \frac{z}{(1+z)^{4/3}} \left(1 + \frac{3g}{5b} - \frac{d}{b} \right)$$

This expression degenerates into eqn. (11) when b tends to infinity, and is useful for evaluating F when g/b and d/b are known small fractions.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

PART C—MONOGRAPHS, MARCH 1957

CONTENTS

	PAGE
Discussion on 'Peak-Voltage Measurements of Standard Impulse Voltage Waves'	1
An Investigation of the Eddy-Current Anomaly in a Low-Silicon Sheet Steel	2
The Measurement of Magnetron Frequency Pulling	8
A Method for the Approximate Determination of the Impulse Response of a Number of Identical Circuits in Cascade	13
The Use of Lossy Material to Suppress Unwanted Modes in Cavity Resonators	25
Hall Effect and its Counterpart, Radiation Pressure, in Microwave Power Measurement	35
Radio Propagation over a Discontinuity in the Earth's Electrical Properties—I	43
Characteristics of the Trigatron Spark-Gap	54
Synthesis of Transmission Systems in Terms of Tandem-Connected Quadripoles	62
A New Form of Electrolytic Tank	81
The Diffraction of an Electromagnetic Wave by a Circular Aperture	87
Theory and Equivalent Circuits of the Double Induction Regulator	96
Representation of Saliency on A.C. Network Analysers	108
The Approximate Solution of Electric-Field Problems with the Aid of Curvilinear Nets	116
Governing in Power Systems by Time-Error	130
Radio Propagation over a Discontinuity in the Earth's Electrical Properties—II. Coastal Refraction	139
Strain-Energy Calculations in the Design of Cat's Whiskers for Semi-Conductor Devices	148
The Calculation of Cyclic Rating Factors for Cables Laid Direct or in Ducts	154
The Production of a Sinusoidal Flux Wave, with particular reference to the Inductor Alternator	167
Properties of Synthetic Quartz Oscillator Crystals	174
The Effect of Magnetic Saturation on the D.C. Dynamic Braking Characteristics of A.C. Motors	185
The Describing-Function Analysis of a Non-Linear Servo Mechanism subjected to Stochastic Signals and Noise	193
The Change of Air-Gap Flux in Electrical Machines due to the Displacement of Opposed Slots	204
Non-Linear Distortion in Transistor Amplifiers at Low Signal Levels and Low Frequencies	208
Intermodulation in Bridge Detector Amplifiers	217
On the Amplification Factor of the Triode	222

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) which are out of print and from which reprints are not available. The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.
Example.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

The Benevolent Fund



Have YOU yet responded to the appeal for contributions to the

HOMES FUND

The Court of Governors hope that every member will contribute to this worthy object

Contributions may be sent by post to

THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF
ELECTRICAL ENGINEERS, SAVOY PLACE, LONDON, W.C.2

or may be handed to one of the Local Hon. Treasurers of the Fund



Local Hon. Treasurers of the Fund:

EAST MIDLAND CENTRE R. C. Woods
IRISH BRANCH A. Harkin, M.E.
MERSEY AND NORTH WALES CENTRE D. A. Picken
NORTH-EASTERN CENTRE J. F. Skipsey, B.Sc.
NORTH MIDLAND CENTRE J. R. Rylands, M.Sc.
SHEFFIELD SUB-CENTRE F. Seddon
NORTH-WESTERN CENTRE W. E. Swale
NORTH LANCASHIRE SUB-CENTRE
G. K. Alston, B.Sc.(Eng.)

NORTHERN IRELAND CENTRE G. H. Moir, J.P.
SCOTTISH CENTRE R. H. Dean, B.Sc.Tech.
NORTH SCOTLAND SUB-CENTRE P. Philip
SOUTH MIDLAND CENTRE Capt. J. H. Patterson, R.A.
RUGBY SUB-CENTRE P. G. Ross, B.Sc.
SOUTHERN CENTRE G. D. Arden
WESTERN CENTRE (BRISTOL) A. H. McQueen
WESTERN CENTRE (CARDIFF) David J. Thomas
WEST WALES (SWANSEA) SUB-CENTRE O. J. Mayo
SOUTH-WESTERN SUB-CENTRE W. E. Johnson

THE BENEVOLENT FUND

Published by The Institution, Savoy Place, London, W.C.2.

Telephone: Temple Bar 7676.

Printed by Unwin Brothers Limited, Woking and London.

Telegrams: 'Voltampere, Phone, London.'